

nature

MOON-BASED
ASTRONOMY

A liquid mirror
is good as silver

LIQUID BIOFUELS

Cutting down
on the alcohol

GENOMIC

INSTABILITY

A mouse model
of human cancer



HOW GREEN IS THIS?

Formula 1 in race for fuel economy

NATURE INDEX

Building a career

Asia on the rise

The balance of scientific power is moving east as scientists in the Asia-Pacific region learn to collaborate more effectively.

Earlier this month, the Tokyo-based Asia-Pacific arm of Nature Publishing Group, which publishes this journal, celebrated its twentieth anniversary by assembling 200 of the region's leading scientists to discuss scientific networking.

Ryoji Noyori, president of RIKEN, Japan's largest research institute, who won the 2001 Nobel Prize in Chemistry for his work on catalysing certain forms of chiral molecules, nicely summed up the thrust of the event: "Asians must be aware that there should be three players: America, Europe and Asia."

There can be no doubt that this mentality — the sense that Asian science exists as a meaningful entity, absorbing contributions not just from Japan, China, Korea and southeast Asia, but also from India, Australia and elsewhere — is slowly gaining acceptance.

But what shape will a more unified concept of Asian science eventually take? It will clearly be a mammoth enterprise, capturing the scientific aspirations of half of the world's population. It will coexist with fraught political relationships: Japan, China and South Korea, for example, remain in conflict over interpretations of responsibility for past wars and ownership of territory. Taiwanese scientists have difficulty travelling to the Chinese mainland and are sometimes left out of scientific meetings. Singapore and Malaysia are often at loggerheads over water supply. The list goes on; by comparison, disagreements between Europe's scientific powers are mere schoolyard spats.

An Asian research community worthy of the name would be a formidable entity. And Asia-Pacific researchers have steadily increased their share of the scientific papers monitored by ISI, the Philadelphia-based citation service, from fewer than 15% in 1990 to 25% in 2006.

They are particularly strong in the physical sciences. In terms of the raw volume of physics papers published in ISI-tracked journals, for example, the region is neck-and-neck with the United States, and slightly behind Europe. In chemistry, it is level with Europe and far ahead of the United States. In engineering, all three regions are in a dead heat, and in photonics, Asia has a commanding lead. In nanoscience and materials, the region accounts for nearly as much as the United States and Europe combined. It is true that the average paper from the Asia-Pacific region is cited less often in the literature — but this performance gap is shrinking rapidly.

Changes in Nature Publishing Group's presence in the region reflect this growth. There were four employees in its Tokyo offices in 1990; today there are more than 50 in a region that accounts for about one in six of *Nature*'s global subscriptions. *Nature Photonics* has its editor-in-chief based in Tokyo and *Nature Nanotechnology* has a manuscript editor there. Among many regional activities, the *Nature China* website, launched this month, serves to highlight the nation's hottest research findings. All this expansion reflects the growing role of the region in the world of science.

But that growth places an onus on Asian researchers to look on each other as viable collaborators — something they have not always done

in the past. Edison Liu, head of the Genome Institute of Singapore, told the anniversary forum that scientists in the region often maintain a 'West-first' approach to collaboration.

There are several deep-seated reasons for this, including strong academic ties to places where the scientists were often trained, in Europe and North America. Western funding bodies, such as the US National Institutes of Health, the Pasteur Institute and the Wellcome Trust, have been more international in their outlook than their peers in Asia have managed to be.

The customary allure of prestige also plays a role: "A collaboration with Boston is more likely to get you noticed by your chair" than one with a partner closer at hand, Liu says. And researchers in Beijing are more likely to be aware of what is happening in Boston than in Tokyo or, for that matter, in Guangzhou.

One proactive step that many scientists in the region could take would be to make themselves more visible on the Internet. Many still don't have their own homepages, and for those that do, basic contact information and research descriptions are often missing or out of date. This is true even of scientists in Japan and Korea — two of the most 'wired' nations on Earth.

The problem is sometimes exacerbated by the commonness of some family names. Around one-fifth of Koreans have the last name Kim, for example. In recognition of the same problem in China, where 1.3 billion people carry only about 100 family names, the government is considering measures to get people to adopt the practice of taking their mother's name as well as their father's.

On a larger scale, researchers need to work hard to circumvent political tensions and establish regional institutions of genuine quality. Some are already emerging, such as the Asia-Pacific International Molecular Biology Network. But according to one survey of global scientists, far fewer have heard of it than of the European Molecular Biology Organization.

There have already been collaborative successes: the Pan-Asian SNP Initiative, for example, brought together scientists from 11 countries, with diverse resource levels, to analyse genome variations and map out human migration patterns across the continent.

Other areas of common interest, such as global warming, will serve to pull Asian scientists more closely together. At a meeting of the science ministers of South Korea, Japan and China in Seoul in January, energy, bioinformatics, disaster prevention and mitigation, and traditional medicine were identified as holding promise for collaboration.

If these three countries can manage to pull together, they will serve as a powerful anchor for collaboration throughout the wider region. This is certain to broaden and deepen in future, benefiting researchers who have much to gain from getting to know their neighbours. ■

"Areas of common interest, such as global warming, will serve to pull Asian scientists more closely together."

Discourse with Iran

Academic freedom is under threat, again.

IShould you wake up one day to find your wife or child or parent in the hands of the secret police in a country that routinely violates the rule of law, you will likely choose quiet probing over publicity. You have no recourse to law or courts. You fear publicity may make things worse. You believe, almost always wrongly, that if you work quietly, use the contacts you have and wait reasonably, the nightmare will be over.

In this plea in the *Los Angeles Times*, Shaul Bakhash, a specialist in Middle Eastern history at George Mason University in Fairfax, Virginia, eloquently captures the predicament of prisoners and their families caught up in political conflicts. Bakhash's wife Haleh Esfandiari, director of the Middle East Program at the Woodrow Wilson International Center for Scholars in Washington DC, is one of three US–Iranian researchers jailed in Iran on charges of plotting a 'velvet revolution'.

Bakhash has broken his silence, and is backing protests by academic and human-rights organizations as the best hope for justice for the jailed researchers (see page 890). Scientists should support such efforts in every way they can, to ensure that Iran's rulers know the world is watching.

Keeping the cases in the spotlight is also helpful in pressing diplomats to make resolution of the cases a priority. The importance of this has been demonstrated in the case of six health workers condemned to death in Libya. Their appeal will be judged today by Libya's

Supreme Court — one step in a delicate endgame, which observers hope will lead to their prompt release.

In the current climate of mutual suspicion between the United States and Iran, simply protesting innocence is not enough. Academics should also demand that Iran respect its commitments to human-rights treaties by making public its evidence against the three, as it has so far failed to do, and allowing them access to lawyers.

Iranian academics are also suffering discrimination abroad as a result of Iran's stand-off with the international community and in particular with the United States. Although the difficulties of getting visas for entry to the United States have eased, Iranians tell of new problems in Canada and Australia, and complain of being shunned in international collaborations as being part of a 'rogue state'.

The arrests have led to calls for a deepening of Iran's isolation (to be fair, these calls are based more on legitimate concerns about safety of travel than on a desire to boycott Iran). This approach has also reared its head in Britain, where the annual meeting of the University and College Union voted on 30 May to ask its branches to consider a proposal to boycott Israeli academic institutions. They should firmly reject this proposal (see *Nature* 417, 1; 2002).

Where colleagues suffer as a result of political tensions, researchers everywhere should be engaging more, not less, in constructive reform and cooperation — through increased support, for example, of collaborative projects in the regions concerned. ■

"Scientists should support such efforts in every way they can, to ensure that Iran's rulers know the world is watching."

An end in sight

Better days ahead for flagship regulator.

One of the great paradoxes of American life is that of the nation's self-image as the "land of the free" and the reality of living in one of world's more tightly regulated societies.

If John Wayne were to stroll down certain Main Streets at noon today, he would not just risk being arrested for jay-walking and carrying an exposed weapon; he might also face charges for smoking tobacco in the saloon the previous evening, or lighting a bonfire in his own back garden. And quite right, too. Sensible regulation has become an integral part of the American way of life.

The European Union may talk a good talk when it comes to rules and regulations, but the truth is that effective measures to clean up water and air — to give just two examples — were pioneered in the United States and are tirelessly enforced there by federal agencies of formidable power and reach.

The largest and most influential of these is the Environmental Protection Agency (EPA), which was founded by President Richard Nixon in 1970 during the environmental movement's first, noisy spring (see page 892). Today, the EPA is an US\$8-billion agency, and its 17,000 staff carry an array of legal and technical expertise, and legislative authority, that sister agencies abroad can barely dream of.

US President George W. Bush and a number of his key supporters

abhor the EPA with an unusual venom. However, it has never been expedient, or even polite, for them to say as much. Furthermore, under a political system renowned for creating budgetary deficits, it has not been feasible for them to actually cut back the agency to any significant extent.

Instead, since 2000, parts of the EPA have existed in a peculiar limbo. Inside the ironically named Ronald Reagan Building and other agency premises, lawyers in some sections are paid handsome wages to do very little, rather than pursue major regulatory infringements that — they know full well — their politically appointed bosses will not follow through in court. 'Commissioners' from the White House Office of Management and Budget or the Council on Environmental Quality roam the corridors, making sure no one is getting too zealous, and reporting back to regulated industries on things to dodge or cover up.

Thankfully, these creepy characters lack the authority to actually fire the conscientious lawyers and scientists who staff the EPA. Like their colleagues at other regulatory agencies, notably the Food and Drug Administration (where political interference, although present, has been less brazen), these people are lying low, aware that no future administration, Republican or Democrat, is likely to hold the agency's underlying mission in such contempt. With a more supportive Congress, and a Supreme Court asserting the agency's power to regulate carbon emissions, an unfortunate period in the EPA's history is drawing to a close. ■

RESEARCH HIGHLIGHTS

Arctic travel

Science 316, 1606–1609 (2007)

Plant species driven north by climate change shouldn't have too much trouble making the trip, according to a survey of plants growing on the rocky Arctic outpost of Svalbard, Norway. Seeds from all of the species studied travelled there from places as distant and diverse as Greenland, Scandinavia and Russia.

Researchers led by Inger Greve Alsos, now at the University Centre in Svalbard, compared genetic samples from nine Svalbard species with those from the same species gathered at sites across the Northern Hemisphere. They found that the island had been colonized by plants from each of the sample locations, presumably travelling by wind or on drifting ice or wood.



B.E. SANDBAKK

BIOTECHNOLOGY

RNA en masse

Nature Methods doi:10.1038/nmeth1058 (2007)

Scientists in France have developed the first efficient method to generate high-quality RNA samples for use as research tools.

Although the importance of diverse RNA molecules in most cellular processes is becoming ever clearer, the methods for producing pure RNAs have remained laborious and costly.

The simple and reliable technique thought up by Luc Ponchon and Frédéric Dardel from the University of Paris Descartes is analogous to the standard method for making genetically engineered proteins — both exploit the cellular machinery of the bacterium *Escherichia coli*.

In a twist to protect the RNAs from destruction by bacterial enzymes during synthesis, they use a bacterial molecule called tRNA as both a shield and a scaffold.

LOW-TEMPERATURE PHYSICS

Electron movies

Low Temp. Phys. doi:10.1007/s10909-007-9373-2 (2007)

Single electrons have been immortalized on the silver screen.

When a single electron finds itself in liquid helium it creates a bubble, thanks to the mutual repulsion between it and helium's two electrons. Wei Guo and Humphrey Maris from Brown University in Providence, Rhode Island, used sound waves to create pressure within such bubbles, pumping them up to the size of a small grain of dust. The low surface tension

of superfluid helium allowed the electron-encapsulating bubbles to grow to this size before they popped.

The bursting bubbles were big enough to be caught on camera (pictured below), and sound pulses were used to track the motion of an electron after the rupture of its bubble. For each electron, bubbles grew and popped continuously as the sound wave pulsed. The resulting images show most released electrons travelling in a straight line. However, a few electrons shot through the liquid helium in a snake-like manner, possibly following a superfluid vortex.

NANOTECHNOLOGY

Tiniest magnets shape up

Adv. Mater. doi:10.1002/adma.200602374 (2007)

Researchers have engineered several different nanometre-scale crystalline structures from a single magnetic material. Each form, they say, may have a unique application.

Jiao-Ming Qiu and Jian-Ping Wang of the University of Minnesota in Minneapolis built their structures using tiny 'nanomagnets' made of iron and platinum (FePt). By 'sputtering' the FePt molecules onto a surface, they created simple, geometric shapes. At a low sputtering current, they found that FePt nucleated into an icosahedral (20-faced)



structure, whereas at higher currents the particles created different types of octahedron.

The magnetic properties of the larger structures varied, and the authors say they could have a range of uses, from biodetectors to ultracompact hard drives.

NEUROSCIENCE

Touching empathy

Nature Neurosci. doi:10.1038/nn1926 (2007)

'Mirror touch' synesthesia — in which people watching others being touched feel the same touch on their own skin — is a real condition, researchers have found. And, they say, such synaesthetes have more gut-level empathy with others.

Jamie Ward and Michael Banissy at University College London, UK, found that those with the condition make more errors than do control subjects in deciding where they themselves are being touched when they observe another person being touched at the same time.

They also found that mirror-touch synaesthetes show more emotional empathy, but not the type of empathy that requires reasoning. The researchers say the results support the concept that we empathize with others by automatically simulating their experiences.

MICROBIOLOGY

Stuck in the gut

Proc. Natl Acad. Sci. USA 104, 10637–10642 (2007)

Scientists have identified a protein structure that helps the food-borne pathogen *enterohaemorrhagic Escherichia*

coli (0157:H7) stick to the host tissue it is invading.

These bacteria cause life-threatening bloody diarrhoea or kidney damage. But Jorge Girón from the University of Arizona in Tucson and his colleagues showed that the gene encoding this protein, known as *E. coli* common pilus (ECP), is present in almost all of the 169 strains of *E. coli* they studied, including non-pathogenic ones that live peacefully in the gut.

The authors suggest that pathogenic strains may exploit this protein to disguise their pathogenic nature from the host's immune system.

IMMUNOLOGY

Inflammation control

Science doi:10.1126/science.1145697 (2007)
New research has shown that retinoic acid, a metabolite of vitamin A, helps determine the course of T-cell specialization and could be important for regulating inflammation.

Hilde Cheroutre and her colleagues at the La Jolla Institute for Allergy and Immunology in California found that blocking retinoic acid perception in mouse T-cell cultures promoted synthesis of pro-inflammatory cells. Conversely, adding retinoic acid to cultures or live mice lowered the production of pro-inflammatory T-cells.

The researchers also found that inflammation could be suppressed in mice if the animals were given T cells pretreated with both retinoic acid and a protein involved in determining T-cell specialization. The team suggests that therapies targeting retinoic acid may be useful against certain inflammatory disorders.

JOURNAL CLUB

Nicolas Gruber
Swiss Federal Institute of Technology, Zürich, Switzerland

A climate scientist worries that attempts to curb atmospheric carbon dioxide levels are challenged on two fronts.

What is your carbon footprint? I must admit that, as someone who frequently travels across continents, mine is well above the Swiss average. Even worse, my footprint has grown over the past few years despite the fact that I am

well aware of the consequences of my actions.

Now, imagine that everyone else on this planet has increased their carbon footprint as well. This is not hypothetical. A recent paper tells us that global carbon emissions have grown at the unexpectedly high rate of more than 3% per year since 2000 (M. Raupach *et al.* *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0700609104; 2007).

In particular, the rapidly increasing appetite for energy of the emerging markets in Asia has led to a dramatic increase in fossil-fuel burning. As a result,

global CO₂ emissions now exceed the worst-case scenarios of just a few years ago. This is far from the direction that we ought to be taking to achieve a stabilization of greenhouse gases that "prevents dangerous interference with the climate system", as the Climate Convention in Rio set out to achieve in 1992.

Unfortunately, the situation may become even more difficult. Earth's biosphere has so far helped to mitigate the carbon problem by removing a substantial fraction of the emitted CO₂, but this 'sink' function may diminish.

There is some evidence that sinks are already weakening (C. Le Quéré *et al.* *Science* 10.1126/science.1136188; 2007), and coupled climate-carbon-cycle models tend to support the view that the trend will persist. If so, we are challenged at both ends — by unexpectedly rapidly increasing emissions and by diminishing sink strengths — making climate stabilization a truly grand challenge.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

CHEMISTRY

Gold leaf

Langmuir doi:10.1021/la063738o (2007)

Large, smooth films of gold would make the precise measurement of forces on the surfaces of materials easier. But making films big and smooth enough has been a challenge.

Now, Liraz Chai and Jacob Klein of the Weizmann Institute of Science in Rehovot, Israel, describe a method for creating gold



films that are smooth at the molecular level and as big as one square centimetre (pictured below) — several orders of magnitude larger than previous films of this type.

The key is to coat one side of thin, single-crystal mica sheets with evaporated gold, then strip away the mica to leave the film behind.

PLANETARY SCIENCE

Mystery of Mars's methane

Geophys. Rev. Lett. 34, L11202 (2007)

The methane recently observed in the martian atmosphere could have been released from methane hydrates below the planet's surface, a new model suggests.

Methane hydrates are one possible source of the gas, but the temperature and pressure below the planet's frozen surface would theoretically stabilize any stores to a depth of 6 kilometres.

Megan Elwood Madden and her colleagues at Oak Ridge National Laboratory in Tennessee now suggest that increasing salt levels may destabilize the hydrates. Their calculations from some predicted sources of high salinity suggest that hydrates might only be stable down to 1.7 kilometres in high-salinity systems. They say the decrease in hydrate stability with increased salinity might allow methane to be released, and escape through fracture zones into the atmosphere.

CELL BIOLOGY

Stem-cell variation

Nature Biotechnol. doi:10.1038/nbt1318 (2007)

Stem cells around the world have a lot in common but are not exactly the same, reports an international consortium.

The International Stem Cell Initiative analysed 59 human embryonic stem-cell lines held in 17 laboratories worldwide to begin standardizing information about these notoriously finicky and mutable cells.

Participating labs considered characteristics such as surface proteins and the expression of certain genes. They found that the lines share most traits — for instance, all express a particular suite of cell-surface markers. However, they differ in others, such as the means by which one X chromosome is inactivated in female lines.

The results should give scientists more confidence in designing studies and comparing their results, the consortium says.

NEWS

Academic freedom under threat in Iran

The jailing of three US-Iranian academics in Iran on accusations of fomenting a 'velvet revolution' signals that Iranian academics and academic freedom could be casualties of the tense stand-off between Iran and the United States.

The Iranian judiciary has said it will decide this week whether to indict the imprisoned scholars on charges of espionage and threatening national security. Last month, the country's intelligence ministry warned Iranian academics and scientists that contact with foreign institutions, or attendance of international conferences, could result in their being considered spies.

Haleh Esfandiari, the 67-year-old director of the Middle East Program at the Woodrow Wilson International Center for Scholars in Washington DC, and Ali Shakeri, a founder of the University of California's Center for Citizen Peacebuilding, were arrested on 8 May. Kian Tajbakhsh, a researcher working on humanitarian issues with the Open Society Institute, based in New York, was arrested three days later.

The three are being held in Tehran's Evin Prison, long notorious for the torture of political prisoners. Their incarceration has triggered protests from academic and human-rights organizations worldwide, including the American Association for the Advancement of Science, Amnesty International, the American Association of University Professors, Human Rights Watch, and Scholars at Risk.

The broader implications of these developments have not been lost. The arrests "raise grave concerns about the ability of internationally recognized scholars and intellectuals to safely visit Iran," notes Scholars at Risk, an international network of universities and colleges that seeks to defend the human rights of academics around the world. The organization thinks they signal a "wider attempt to intimidate intellectuals and to limit academic freedom in Iran."

The crackdown is already having a "chilling effect" on international collaboration, says Fatemeh Haghigatjoo, an Iranian psychologist and visiting fellow at Harvard's John F. Kennedy School of Government. She likens it to the repression of foreign contacts in the Soviet Union at the height of the Cold War.

For the moment, researchers in the natural sciences seem to be less affected than their

social-science colleagues. Haghigatjoo, a reformist former member of parliament and a renowned human-rights advocate, says academics who deal with Iran's social, cultural or political situation are at the most risk.

Yousef Sobouti, an astrophysicist and director of the Institute for Advanced Studies in Basic Sciences (IASBS) in Zanjan, Iran, seems to confirm as much when he says he has never encountered difficulties with the regime. "Just today, I signed letters for eight of our students and faculty members to attend conferences in countries ranging from Japan, Russia to Europe and the United States." Reza Mansouri, a physicist at Sharif University of Technology in Tehran, and a deputy science minister under reformist former president Mohammad Khatami, is similarly sanguine. He shrugs off the intelligence ministry's threat as a warning — aimed mainly at naive younger researchers — to be vigilant in all dealings with the West.

Mansouri says Iranian academics have long been targeted by foreign intelligence services, and claims such efforts have increased recently. "Intelligence agents have tried hard to contact Iranian scientists visiting research institutions abroad, or attending conferences," he says, "with the very obvious goal of having intelligence information about Iran".

Some stories of such approaches are "unfortunately true," agrees David Rahni, an Iranian-born chemist at Pace University in New York. But Rahni is less cool than physical scientists *Nature* spoke to in Iran about the impact of recent events. He says they will make colleagues in Iran "think twice about communicating; it will have a negative impact on the country's science".

For Haghigatjoo, the regime is rolling back years of slow progress on the opening up to the outside world undertaken by presidents Rafsanjani and Khatami. Of 20 Iranians she invited to a recent workshop on the future of democracy in Iran at the Massachusetts Institute of Technology in Cambridge, only two showed up, she says, with the others declining for fear of retribution or after the Iranian security services forbid them to attend.

But, like other academics, Haghigatjoo dismisses the idea that the crackdown represents a



return to the anti-science, anti-academic years that followed the inception of the current Iranian theocracy in the 1979 revolution. She sees it more as a tactical response to current international and domestic politics.

President Mahmoud Ahmadinejad's government is increasingly unpopular at home, risking defeat in next year's parliamentary elections and in the presidential elections of 2009. The current repression, says Haghigatjoo, is not like the purge of university liberals that followed Ahmadinejad's 2005 election. It is simply that of a regime clinging on to power through "zero tolerance of any dissent" and is targeting any opposition, including that from



**TWIN BROTHERS
MAKE WOMEN LESS
FERTILE**

Testosterone sharing in the womb has knock-on effects.
www.nature.com/news

BANANASTOCK



Torture of Iran's political prisoners is reportedly commonplace at Tehran's notorious Evin Prison.

conservatives, that threatens its grip.

And the United States has made no secret of the fact that it is spending millions to encourage some of the forms of dissent under attack, under the rubric of "pro-democracy support". Haghigatjoo, along with many academic and human-rights groups, sees this as both incomprehensible and counterproductive, playing into the hands of the regime by providing a pretext to attack independent pro-democracy groups and academic reformers.

"How can we as Americans spend millions of dollars to effectuate covert or overt actions

against a sovereign government, and expect them to take that crap sitting down?" says Rahni. Esfandiari's husband, Shaul Bakhsh, himself an academic expert on Iran, has also attacked the policy: "Loose talk of regime change and allocation of money supposed to advance democracy in Iran has done a great deal of harm to Iranian academics, intellectuals and researchers," he told the *Financial Times*. "It also feeds the paranoia of the Iranian regime of American intentions."

Declan Butler

See Editorial, page 886.

Monkey stem cells cloned

CAIRNS

Cloned embryonic stem cells have at last been generated from monkeys, a US research group claimed this week. The work was announced on 18 June in a last-minute presentation at the annual meeting of the International Society for Stem Cell Research in Cairns, Australia. These findings will renew hopes that similar cells can be produced for humans.

"We've been waiting for this for some time," said Alan Trounson of Monash University, Victoria, Australia, who introduced the presentation.

The work was carried out by Shoukhrat Mitalipov of the Oregon National Primate Research Center in Portland and colleagues. They removed the chromosomes from unfertilized monkey eggs and replaced them with nuclei from the skin cells of an adult rhesus monkey (*Macaca mulatta*). A total of 278 oocytes yielded 21 blastocysts (hollow early embryos), from which the team eventually derived two embryonic stem-cell lines. The work has not yet been published.

The failure of earlier attempts to clone embryonic stem cells in this way using monkeys had led several experienced researchers in the field to suggest that characteristics specific to primates might make it impossible (C. Simerly, et al. *Science* 300, 297; 2003). "Now we know primates are possible, like other mammal species," says Norio Nakatsuji from Kyoto University, who has established primate stem-cell lines from uncloned embryos.

One possibly crucial aspect of the new work is a gentler way of removing chromosomes from the egg by using imaging software rather than staining and ultraviolet light to guide the process.

José Cibelli, a cloning expert at the University of Michigan, Ann Arbor, says there is no clear reason why techniques to make human embryonic stem cells through nuclear transfer need be very different from those used in non-human primates. But he cautions that "what works in rhesus monkeys doesn't work in baboons."

The Oregon group's work has yet to be replicated in monkeys, but Renee Reijo Pera at Stanford University, California, plans to apply the techniques to other primates. He says that success in primates will renew the resolve to find similar techniques for humans.

Monya Baker

Monya Baker is news editor at *Nature Reports Stem Cells*.

www.nature.com/stemcells/index.html

SPECIAL REPORT

All the King's men

Congress requires the Environmental Protection Agency to look after the environment; the Bush administration wants it to be pro-business.

Emma Marris steps through the looking glass to profile the result.

In 2006, a federal appeals court was called on to examine the way that the US Environmental Protection Agency (EPA) was proposing to regulate power-plant emissions. Congress had decreed that regulations should be revisited after "any physical change" that increases emissions. The EPA and counsel for various industries argued that if the word "any" was construed properly, old power plants could have their equipment significantly revamped without "any physical change" taking place, and that if Congress had wanted to address such cases it would have spelled things out more clearly.

The court was having none of it: "Only in a Humpty-Dumpty world would Congress be required to use superfluous words while an agency could ignore an expansive word that Congress did use. We decline to adopt such a world-view." The ruling resonated with a great deal of the criticism of the EPA under the George W. Bush administration; many environmentalists hold that if the agency could indeed, like Humpty Dumpty in Lewis Carroll's *Through the Looking Glass*, make words mean what it wanted them to mean, then an awful lot of words would end up meaning "OK, pollute a bit if you must."

Outweighing benefits

The EPA's remit means it will never be a stranger to criticism. A draft report by the White House Office of Management and Budget (OMB) put the annual costs of government regulation over the past ten years at between \$39 billion and \$46 billion in 2001 dollars, around 60% of which was attributable to EPA regulations. As a result, the agency regularly gets flak from those facing those costs, mainly industries. The same report, though, calculates the benefits of regulation at \$98 billion to \$480 billion, with about 85% attributable to the EPA. If the EPA is already doing that much good, environmentalists argue that it could do even better by regulating more stringently.

The agency has been attacked from both sides since it was founded in 1970. But things have been particularly tough and one-sided over the past six years, during which time

the agency has endured a constant, low-level rumbling of scandal. Standards for smog were called deadly and rules for mercury too lenient; enforcement changes a gift to the coal industry before the 2004 elections; and much more.

All this comes as the EPA faces what is arguably the biggest challenge in its history: regulating the main gases that cause climate change. The agency had until recently claimed that, when words meant what it wanted them to mean, carbon dioxide did not count as pollution, and thus did not come under its remit. This April, the Supreme Court told the agency

that that wasn't going to wash. Carbon dioxide now needs to be assessed as a factor in regulating car emissions. The case marked a victory for California governor Arnold Schwarzenegger, among others, who have been keen to regulate along these lines but was stymied by the EPA's stance.

Most of the climate-change bills currently swirling around Washington lay out a cap-and-trade system to restrict greenhouse-gas emissions. And nearly all of them specify the EPA as the lead agency. This is in part because the sulphur-trading scheme that the EPA set up in the 1990s as a market-based approach to controlling acid rain is widely admired.

Under the current administration, though, the emphasis has not been on innovative regulation but on less regulation, with a reliance on voluntary approaches and a close, trusting relationship with the regulated industries. The

EPA head Stephen Johnson (right) has been on the opposite side of arguments from California's Governor Arnold Schwarzenegger on controlling green-house gas emissions.



C. KNORR/DESIGN PICS/CORBIS



Bush administration has a record of pushing its legal authority as far as it can to influence its agencies in their interpretation of statutes, and it has done so vigorously at the EPA; hence the overturns in the courts.

President Bush has been served by three EPA administrators. The first, former governor of New Jersey Christine Todd Whitman, had the highest profile. She resigned in 2003, apparently frustrated with meddling from the White House, where Vice-President Dick Cheney's work on energy policy was hard to square with her environmental agenda. The most recent

W. FOSTER/OFF. GOV. SCHWARZENEGGER/SIPA PRESS/NEWSCOM



ARCTIC SPRING COMES TWO WEEKS EARLY
Plants and animals show big spring-time shift over a decade.
www.nature.com/news

T.T.HOYE SOURCE: SEPA



Rules set by the EPA's air office show greater benefits than those from other federal agencies.

administrator, Stephen Johnson, an agency veteran and the first career scientist to rise to the level of administrator, has kept a lower profile. Russell Train, who served as agency head under the Republican presidents Richard Nixon and Gerald Ford, says he feels that the EPA has "lost its way" under Bush. Train, now at the conservation group WWF in Washington DC, says that in his day the EPA was on a longer leash. "Never ever once was I ever told by anybody — either in the White House or purporting to speak for the White House — how to make a decision or how to reach a decision."

Balanced interests

The White House seems to exert considerable influence over the EPA's decisions. It has, for example, expanded the power of the OMB, which must approve all new regulations and standards. Another White House office, the Council for Environmental Quality (CEQ), coordinates federal environmental policy and can put the kibosh on any agency plans that run foul of other departments' priorities or the president's agenda. James Connaughton, head of the CEQ, says that people not involved in the interagency process don't realize that interests

have to be balanced across the government. "From the pure environmental perspective I would have recommended one approach, but from the economics another approach would be indicated. And that is rational," he says.

One long-term career employee, who asked not to be named because she is still with the agency, describes the growing clout of the White House: "We used to have knock-down fights with [the OMB], but it is not that way anymore," she says. "Our leaders do not fight back. We basically take our marching orders." Insiders point to the Office of Air and Radiation — one of the largest departments in the agency — as a case study of Bush appointees doing as the White House says.

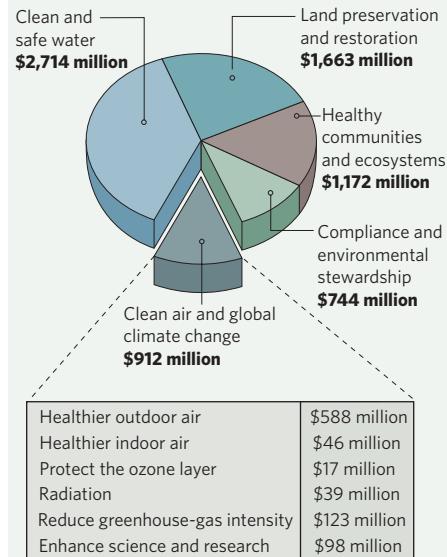
In 2001, Jeff Holmstead was recruited to run the 'air office', as it is known. He previously represented came from Latham & Watkins, a Los Angeles-based law firm that represents chemical and electronics manufacturers, among others. Holmstead and his deputy Bill Wehrum (who had worked for the same firm) were responsible for the Clean Air Interstate Rule, which caps some pollutants from power plants in eastern states, and for the non-road diesel rule, which limits nitrous oxide and soot emissions from construction equipment and the like. Johnson has described these regulations as "two of the five most health-protective clean-air rules in EPA's history". And they have significantly cut down on emissions. The OMB calculates that the air office produces benefits of \$60 billion to \$410 billion for costs of between \$19 billion and \$22 billion.

But Holmstead and Wehrum angered many of the staff when in 2003 they proposed changes that would keep many older power plants from having to adopt modern pollution standards when they remodel. These were the changes eventually over-ruled on the basis of their Humpty-Dumpty approach to language. Bruce Buckheit, head of air enforcement under Holmstead, resigned in protest at the way the rules forestalled planned action against polluters. Now an environmental consultant and sailing instructor in Virginia, he says that the White House was running the show. "Basically Jeff Holmstead and Bill Wehrum would go and meet with Jim Connaughton and come back with their orders," he says.

Independent thinking

Holmstead, who has worked at the law firm Bracewell & Giuliani in Houston since leaving the agency in 2005, denies the charge: "I never had direction from the vice-president's office or the CEQ about what line to take." He adds that he cared more about getting things

EPA spending plan 2008



done than pleasing environmentalists. "Some people say that industry got more of what they wanted under [this administration] than under the Clinton administration — but just because industry promotes something doesn't mean it is wrong."

On Holmstead's departure, Wehrum took over as acting director of the air office and was nominated to take the job on permanently. But in April the White House withdrew the nomination, as the Democrats began sharpening their knives for the hearing at which they could question him before confirming or rejecting his nomination. Wehrum resigned on 1 June.

"We have tried to be as smart as we possibly can."

Speaking to *Nature* in his last week on the job, he said that his goal while in office was to hit that "sweet spot" where industry and environmentalists can both be happy.

"What we have tried to do is to be as smart as we possibly can and get the most bang for the buck," says Wehrum.

In January 2009, it seems fairly certain that a new president, even if a Republican, will appoint a new EPA administrator — someone who will probably have to take on the daunting task of regulating carbon dioxide, and will have a demoralized agency staff with which to do so. But the challenge may well help that morale. When the Supreme Court made its decision this April, champagne corks were heard to pop discreetly in some parts of the agency's Pennsylvania Avenue headquarters. "It has been a real morale booster," says the career employee, looking to the future. "My immediate supervisor says I have got to stay so we can put it all back together." ■

See Editorial, page 886.

Fossils challenge DNA in the dating game

Yet again, molecules and fossils are at odds in the dating of a key event in the history of life. On page 1003 of this issue, palaeontologist John Wible of the Carnegie Museum of Natural History in Pittsburgh, Pennsylvania, and his colleagues use their discovery of a Cretaceous mammal from Mongolia to build the most complete fossil-based mammal family tree so far. It suggests that modern mammals arose only after the dinosaurs went extinct 65 million years ago, thus contradicting dates derived from the DNA of living mammals — including another recent *Nature* paper (O. R. Bininda-Emonds *et al.* 406, 507–512; 2007). In the DNA analyses, modern groups appear earlier, between 100 and 80 million years ago, with the extinction of the dinosaurs seeming to have had little effect on their evolution.

Yawning gaps between molecular and palaeontological approaches to the dating of evolutionary landmarks have appeared ever

since molecular approaches based on DNA sequences first became widely used about 15 years ago. When results differ, the molecular technique almost always pushes events further back than the fossil record. For example, some molecular studies estimate that multicellular animals arose about a billion years ago, but the fossil record goes back a mere 600 million years. Fossil plant spores date back 475 million years, but molecular dates put plants on land 700 million years ago.

However, there are signs that this deep-running discord may be being resolved, at least in some cases. “Ten years ago the divide looked very sharp,” says palaeontologist Mike Benton of the University of Bristol, UK. “But there’s been a certain amount of movement both ways — palaeontological and molecular people are moving towards one another.”

That molecular dates are a little earlier is to be expected. DNA should record the moment that

a lineage split, but the imperfections of the fossil record mean that the first preserved post-split organisms will date from later on. And there are some groups — such as the microscopic, soft-bodied animals that make up about half the extant animal phyla — that have no fossil record at all. “We should be willing to accept biases in the fossil record regarding certain types of animal,” says Blair Hedges of Pennsylvania State University in University Park. Hedges has been among the most vigorous advocates of earlier DNA dates for major evolutionary splits, including those that produced the animals, the land plants, the fungi and mammals.

But although Kevin Peterson of Dartmouth College in Hanover, New Hampshire, agrees that some groups’ fossil records are too patchy to be trusted, he is scathing of the notion that the first animals could have gone unfossilized for nearly half a billion years. “To say that because the record is incomplete you can have a

Quarter of deaths from environment are avoidable

Living in an unhealthy environment kills many times more people than die in road accidents, violent conflicts and natural disasters put together, but these risks rarely make headlines. Now an analysis by the World Health Organization (WHO) reveals that about a quarter of the annual health and death toll of living in such environments could be avoided.

The report is the result of an eight-year meta-analysis of relevant scientific literature and available health and population statistics, which went through the fraction of mortality and illness attributable to environmental causes factor by factor. More than 100 public-health experts and epidemiologists advised on the regional health impacts.

Polluted water, poor sanitation and smoke inhalation resulting from indoor use of wood-burning stoves are the main risks in most low-income countries, and poverty is a major obstacle when it comes to reducing risk exposure. But environmental factors pose an avoidable health risk in rich countries as well, says Annette Prüss, a

WHO public-health scientist who oversees the project. Noise, work stress and outdoor pollution all add to the burden of ill health attributable to the environment.

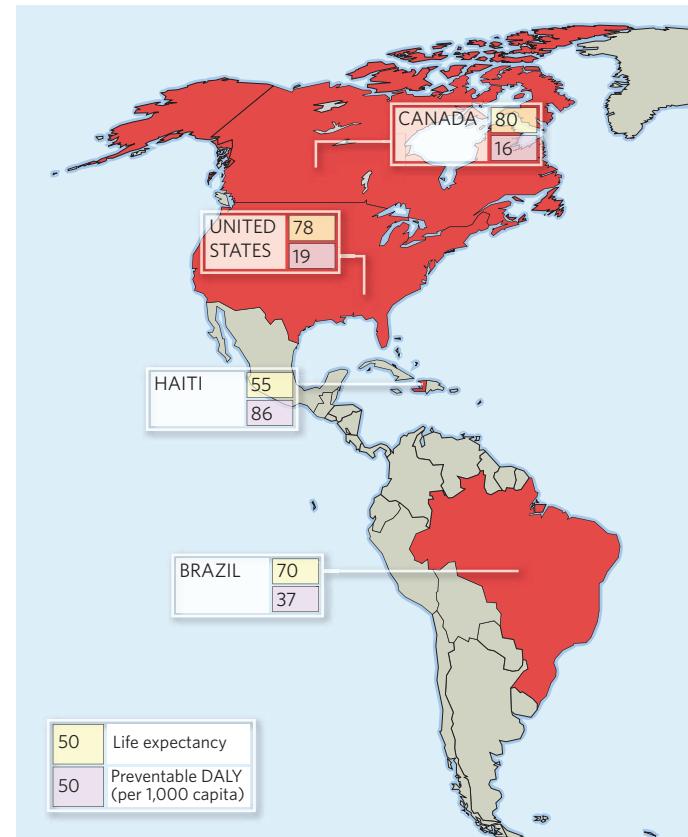
The WHO’s research centres on ‘disability adjusted life years’ (DALY) that are preventable through healthier environments. The DALY is a commonly used unit that includes years lost when someone dies prematurely and also takes account of years blighted by chronic disease or disability.

The data, says Prüss, are to help national health authorities prioritize disease prevention efforts. “Most countries, rich and poor, haven’t yet realized that severe health problems often have quite trivial causes,” she says.

“The WHO exercise is helpful in that it draws attention to the fact that many health problems need a non-medical response,” says Andrew Scott, policy director with Practical Action, a non-governmental organization that helps modernize cooking stoves in Africa. “Small amounts can make a large difference.” ■

Quirin Schiermeier

“Most countries haven’t yet realized that severe health problems often have quite trivial causes.”



billion-year-old bilaterian is a woeful misinterpretation of the fossil record," he says.

DNA clocks have their own problems, which can cause overestimates of the age of evolutionary splits. Different genes can evolve at different rates, and the calibration of DNA clocks must be handled carefully. A calibration that involves a slowly evolving lineage can throw a whole tree off.

Benton sees hope in new statistical methods that can better account for uncertainty in both the timing of calibration points and measurements of evolutionary rates. In such treatments, both fossil and molecular dates go from being fixed points with error bars to being a range of probabilities. "The flexibility is hugely liberating," says Benton. "I see a glowing future for this."

Using such an approach, Peterson has obtained molecular dates for the origin of the animals that match their first appearance in the fossil record (K. J. Peterson & N. J. Butterfield, *Proc. Natl. Acad. Sci. USA* **102**, 9547–9552; 2005). He presented his latest analyses this

**"The flexibility is
hugely liberating —
I see a glowing future
for this."**

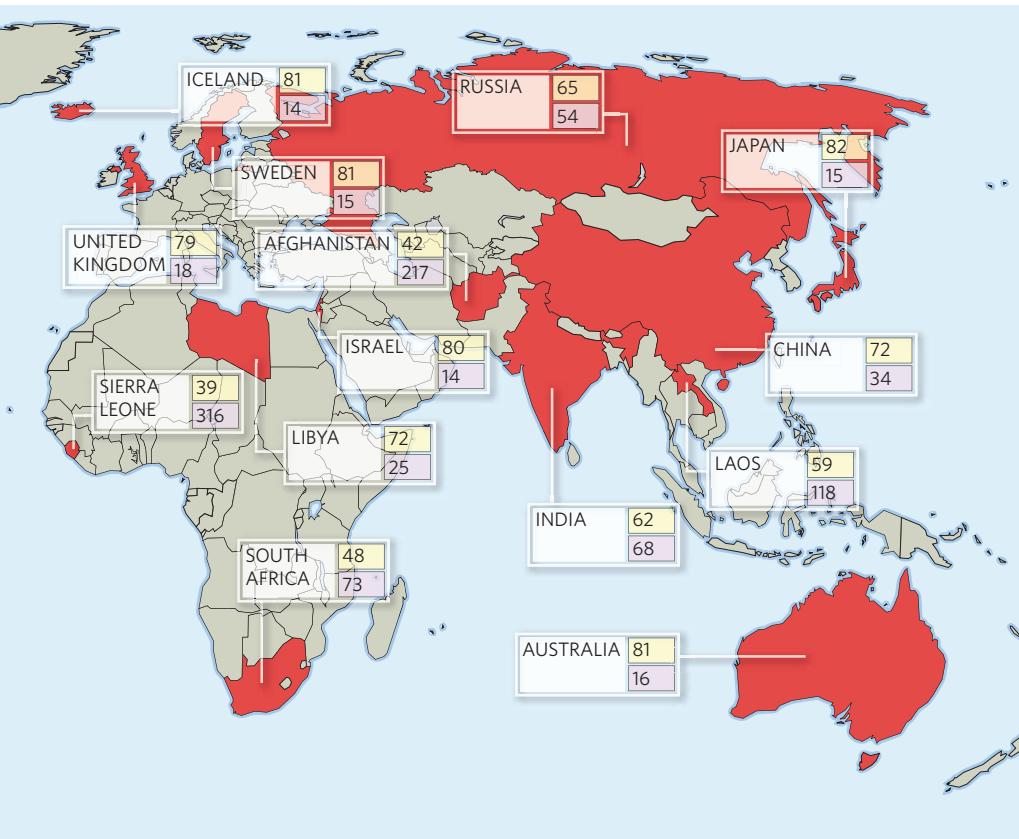
week at a meeting at the Royal Society in London. "All of the latest estimates are putting the division of the bilateria right in the Ediacaran," he says, referring to the geological period predating the Cambrian explosion of about 540 million years ago. A group led by Emmanuel Douzery at the University of Montpellier in

France has got similar results in the past few years. Hedges argues, however, that these more recent dates are the result of calibration errors.

The mammals look to be a tougher nut to crack. "I don't have a good answer as to why there's this discrepancy," says Wible, "but I doubt our work is the final answer." He is part of a project funded by the US National Science Foundation to combine molecular and morphological data to build the most complete mammal tree so far. "We need to marry the data sets together," he says. "Only this type of analysis can address these major problems." ■

John Whitfield

See News & Views, page 918.



ZOO NEWS

Big and cold

Directors of the Alaska Zoo in Anchorage have finally agreed to move Maggie, the world's most northerly elephant, to a new, more comfortable home further south.

A. GRILLO/AP



Small and hungry

Researchers at Japan's Tohoku University in Sendai have proved that it's not just dogs that can perform pavlovian tricks. They have trained their 'Pavlov's cockroaches' to salivate when presented with a specific non-food odour.

SCORECARD

Baby monitors
Chicago science teacher and mum-of-two Natalie Meilinger has been keeping one eye on developments aboard the shuttle Atlantis...after her baby monitor started picking up the latest images from NASA.

Kids with hay fever
A UK study suggests that school kids suffering the summer affliction are twice as likely to drop a grade in their exams, relative to their performance in the winter 'mock' tests.

ON THE RECORD

"Biking is one of the manliest pastimes on the planet, so it's bloody reassuring to know someone's taking care of our virility while we're on the road."

Round-the-world motorcyclist Nick Sanders welcomes Swiss scientists' invention of a coating for black leather trousers that helps them to reflect heat and avoid causing fertility problems for male bikers.

Sources: Reuters, PLoS One, Associated Press, The Scotsman, Revolver Communications

Darwin down but not out

The British government has withdrawn its bid to have Charles Darwin's home and the surrounding countryside designated a World Heritage Site, at the same time expressing strong concern over the way in which sites of scientific heritage are judged.

The decision follows an unfavourable evaluation of the Darwin bid by the International Council on Monuments and Sites (ICOMOS), which advises UNESCO's World Heritage Committee on aspects of cultural heritage. The government says that the evaluation, which has not been made public, questioned the integrity, authenticity and value of the proposed site.

In an open letter to the director of the World Heritage Committee, the UK's Department of Culture, Media and Sport described this assessment as "fundamentally flawed" with "serious weaknesses and omissions". But proposed sites have only one chance with the committee. Fearing the ICOMOS report might sway the decision, the government chose not to let the Darwin nomination go before the committee, which meets on 23 June in Christchurch, New Zealand, to decide on new sites (see 'Natural wonders'). A new Darwin proposal will be submitted in two years' time.

This year's 'Darwin at Downe' proposal sought to put a World Heritage ring around 1,000 hectares of countryside in and around the village of Downe on the outskirts of London. At its heart is Down House, Darwin's home from 1842 until his death in 1882. The surrounding landscape fed Darwin's mind as he wrote *The Origin of Species* and later works: he studied orchids at Downe Bank, found insectivorous plants on Keston Common and collected seeds for his experiments at Cudham School pond.

"I can't think of anything more important to do for the history of nineteenth-century science



Darwin's home won't be put forward for World Heritage listing.

than to protect the whole environment. Darwin inhabited and exploited," says James Moore, a Darwin scholar at the Open University in Milton Keynes and one of the first historians to explore the importance of this rural refuge to Darwin. "Muslims go to Mecca, Christians go to Jerusalem, Darwinians go to Downe," he says.

But without natural wonders or spectacular architecture, Darwin at Downe does not tick



obvious World Heritage boxes. Although he was surprised to hear of Downe's difficulties, Geoffrey Belcher, site coordinator for the Maritime Greenwich World Heritage Site in London, thinks that "A site with a limited range of qualities will be at a disadvantage." The inclusion of the Royal Observatory makes Greenwich one of the few World Heritage Sites to celebrate science, but the site boasts architectural splendour and naval history too.

"This doesn't look like other World Heritage sites," admits Randal Keynes, Darwin's great-great grandson and author of the Darwin at Downe nomination. But he defends its value as testimony to "the understanding of the natural world by observation, hypothesis, experiment, free and wide exchange of information and ideas, theory-building and communication."

Values in question

In 2005 the World Heritage Committee called for "nominations which recognize and celebrate achievements in science", which seemed to make the Darwin at Downe proposal timely. But along with questioning the integrity of the landscape and the authenticity of the restoration of the house and gardens, the ICOMOS report argued that Downe and its surroundings are not of "outstanding universal value" as defined under the World Heritage Convention.

In its letter, the British government makes the broader point that few, if any, scientific sites will be accepted onto the World Heritage List unless a dedicated set of guidelines for their treatment is drawn up. It offers "to host and fund, on behalf of UNESCO, an international expert meeting to examine the issues of representing science on the World Heritage List." ICOMOS declined to comment on the letter.

Despite the setback, English Heritage, the government agency that has owned the Down House estate for the past decade, said it will go ahead with a new permanent exhibition, an education suite for school groups and a resources room where visitors can search the Down House collections, look at Darwin's *Beagle* notebooks online and learn more about his experiments. Culture minister David Lammy says that the government looks forward to submitting a revised bid in 2009, the bicentenary of Darwin's birth and the 150th anniversary of *The Origin of Species*. ■

Henry Nicholls

M. CRABTREE/TROIKA

Natural wonders

Few cultural sites of scientific renown achieve World Heritage status, but natural wonders of interest to science do a bit better. The World Conservation Union (IUCN), which advises the World Heritage Committee on natural sites that might merit its consideration, is recommending that the following be added to the World Heritage list at the committee's forthcoming meeting in Christchurch, New Zealand:

- Primaeva beech forests in the Carpathian mountains, in Slovakia and the Ukraine
- The South China karst, a distinctive rock formation

- Jeju volcanic island and associated lava tubes, off the coast of South Korea
- The rainforests of Atsinanana, Madagascar, which contain much of the island's remaining biodiversity
- Teide National Park, Tenerife, Spain, which includes the peak of the world's third-tallest volcano (and an astronomical observatory)
- Lope-Okanda, Gabon, a landscape boundary between tropical rainforest and relic savannah environments

The IUCN also recommends that the World Heritage Site centred around the Jungfrau in the Swiss Alps be expanded.

A.W.

US universities promise to go carbon neutral

In the biggest move yet to make academia carbon neutral, the leaders of more than 290 US colleges and universities have signed a pledge to reduce or offset all their greenhouse-gas emissions.

Signatories of the American College and University Presidents Climate Commitment range from large public university systems — such as the University of California — to small community colleges. All have agreed to make their plans to go carbon neutral, and their progress in achieving it, available for public scrutiny.

But it's not clear how long the universities will take to reach their goal or how much it will cost. Many are hoping that short-term savings, such as cuts in electricity bills, will help pay for longer-term changes, such as altering buildings to make them more energy efficient or buying carbon offsets to compensate for air travel.

China looks for alternative biofuel options

China is likely to stop growing food crops such as corn (maize) to make biofuels, according to a statement earlier this month by the National Development and Reform Commission.

The policy comes as demand for corn for biofuel is jumping, and prices worldwide are rising. Earlier this month, the United Nations Food and Agriculture Organization reported that worldwide import bills for coarse grains and vegetable oils — used in biofuels — are estimated to rise by up to 13% between 2006 and 2007.

In an attempt to reduce harmful emissions and decrease its need for imported oil, China plans to have 15% of its fuel coming from renewable sources, such as biofuels, by 2020. In place of corn, the China Oil and Food Corporation says it will focus on sorghum to make ethanol.



China intends to switch from growing corn to sorghum for biofuels.

More choice for space tourists

An established European space-equipment company has joined dedicated space tourism ventures such as Virgin Galactic and SpaceX in the race to take passengers to space.

Astrium, a subsidiary of the European Aeronautic Defence and Space Company, has unveiled plans for a vehicle capable of shooting four people to an altitude of 100 kilometres, giving them a 3-minute zero-gravity experience. The €1 billion (US\$1.3 billion) needed for the project is expected to come from private capital, topped up with loans and regional development funding. Astrium hopes to take its first customers up in 2012. Virgin Galactic claims it will be ready for business in 2009.

The Astrium craft will use standard jet



engines to take off and land. Once it reaches an altitude of 12 kilometres, rockets will blast the vehicle the rest of the way.

US genomics centre settles lawsuit with Icelandic firm

A new US genomics centre set up to find inherited diseases in children can continue operations, now that it has settled a potentially crippling lawsuit filed by the Icelandic firm deCODE Genetics.

Last year, deCODE sued the Center for Applied Genomics at the Children's Hospital of Philadelphia in Pennsylvania, saying that four researchers who left the firm took proprietary data with them to start the \$39-million non-profit institute. The centre plans to analyse DNA from 100,000 children.

But on 14 June, a US federal court approved a confidential settlement whereby the four researchers can conduct projects at the new centre. Both parties refused to provide details of the settlement.

Eris, not Pluto, is most massive dwarf planet

Pluto, step aside yet again. The dwarf planet known as Eris is officially the most massive dwarf planet known — heavier than the much-maligned Pluto, which the International Astronomical Union tossed out as a fully-fledged planet last year.

Michael Brown and Emily Schaller, of the California Institute of Technology in Pasadena, have now measured the mass of Eris, known informally as Xena, whose discovery was first reported in 2005. It clocks in at 16.6 billion trillion kilograms, making it 27% more massive than Pluto (M. E. Brown & E. L. Schaller *Science* 316, 1585; 2007).

These measurements were derived from new observations, made by both the Hubble Space Telescope and the Keck Observatory in Hawaii, of the orbit of Eris's moon Dysnomia. Brown and Schaller first worked out that the moon takes about 16 days to

orbit Eris. Kepler's laws of planetary motion and models detailing the gravitational pull between two objects then allowed them to calculate the planet's mass.

Boost for physical sciences in US research budget

Round one of this year's budget cycle in the United States, the first under the new Democratic-led Congress, bodes well for research.

Over the past few weeks, budget committees in the House of Representatives have given a tentative nod towards substantial increases in 2008 for several major science agencies. The Department of Energy's office of science would receive a 19% increase to US\$4.5 billion, and the National Science Foundation would see a 10% boost to \$6.5 billion.

Not all agencies will enjoy such impressive growth. The troubled research wing of the Department of Homeland Security would grow by 3% after being slashed last year to \$630 million, and the National Institutes of Health would grow by 2.6% to \$29.6 billion, which advocates of biomedical research say is not enough.

The House process is far from complete, and bills still need to be drafted in the Senate and eventually signed into law by the president. But the signs look good for research, says analyst Kei Koizumi of the American Association for the Advancement of Science in Washington DC. "It's great news," he says. "More for everyone."

Correction

In the article "Simple switch turns cells embryonic", Shinya Yamanaka is quoted as saying he has not worked with embryos before. In fact, Yamanaka used fibroblasts derived from mouse embryos for the experiments described. In other experiments outlined in the article, the same results were achieved with fibroblasts taken from postnatal skin cells.

BUSINESS

Anaemic outlook for Amgen

A rash of problems has knocked some of the shine off one of the world's top biotechnology companies, as **Meredith Wadman** reports.

April may have been T. S. Eliot's cruellest month, but for Amgen, it was nothing to May. Last month, the company whose name is synonymous with biotechnology success took an unprecedented battering.

The setbacks centred on safety and pricing concerns about Epothen and Aranesp, Amgen's prized pair of anti-anaemia drugs. And they had analysts predicting that the glory days of the Thousand Oaks, California, company in that lucrative market could be gone for good.

"Amgen is a company that's facing significant hurdles at the moment," says Peter Knight, an analyst with Wood Mackenzie, a life-sciences consulting firm based in Edinburgh, UK. However, he adds: "It's easy to paint a very bleak picture and that's probably not very accurate," arguing that Amgen will continue to dominate anaemia drug treatment in the United States.

Others are less sanguine. "They can't even think about the anti-anaemia market coming back," says William Tanner, an analyst at Leerink Swann, an investment bank based in Boston, Massachusetts. "It's going to be a ghost of its former self."

Epothen and Aranesp generated \$6.6 billion of Amgen's \$14.3-billion global revenue last year, while the company's drug pipeline is somewhat anaemic itself, with only one potential blockbuster, osteoporosis medication denosumab, in late-stage trials. "The great thing about denosumab is that it is arguably the best asset in the collective biotechnology-pharmaceutical pipeline," says Mark Schoenebaum of Bear Stearns, a New York broker. "The bad thing is it's all that Amgen has in its late-stage pipeline right now of commercial significance."

Amgen is seeking to rectify that. Earlier this month, it paid \$300 million to buy Alantos Pharmaceuticals, a diabetes-drug company based in Cambridge, Massachusetts, and \$420 million for Ilypsa, a kidney-drug company in Santa Clara, California, whose leading drug candidate would complement Amgen's anaemia remedies in patients with kidney failure.

The two anaemia drugs have all but defined Amgen — and the anaemia market — since Epothen became the company's first marketed drug in 1989. Aranesp, a longer-lasting version of the same genetically engineered, blood-boosting protein, went on sale in 2001, aimed



Under siege: problems mount at Thousand Oaks.

at anaemia caused by cancer chemotherapy. Its sales in particular have driven Amgen's growth since, leaping by 26% in 2006, when total company sales grew by 15%.

But in January, the company announced that cancer patients receiving Aranesp in a clinical trial were more likely to die than those on placebo. Then in February, it emerged that a Danish trial of Aranesp in people receiving radiation for head and neck cancer had been stopped after more deaths and cancer recurrences occurred in the Aranesp-treated group. Amgen had not publicized this news, and the matter is now being investigated by the Securities and Exchange Commission.



The US Food and Drug Administration (FDA) took action in March, slapping a 'black box' warning on both drugs, urging that they be prescribed at the lowest possible doses. On 10 May, a committee of FDA advisors went further, recommending limits on Aranesp's use in cancer patients. The FDA is not bound to follow the advice of its external advisors, but it usually does.

Last month's troubles worsened on 14 May, when Medicare, the US government health programme for the elderly, proposed severely limiting reimbursement for the anaemia drugs when used in cancer. This proposal — which may still be modified — provoked loud complaints from oncologists and Amgen's retort that it is "not supported by scientific evidence". The company has a huge amount riding on the final Medicare ruling, which is expected by September — especially since private insurers often follow Medicare's lead.

Then it emerged that the New York state attorney-general had subpoenaed the company, demanding documents on its activities in drug marketing, pricing and promotion. Amgen received the subpoena the day after *The New York Times* reported that some physicians have profited richly by prescribing Amgen's drugs to Medicare patients.

The damage to sales of Amgen's two blockbuster anaemia drugs cannot be gauged until July, when the company will release its second-quarter results. Analysts have downgraded the stock, however, and the company lost \$9 billion in market value during May, on top of similar losses earlier in the year (see graph).

Company officials declined to comment in detail, instead reiterating public statements emphasizing Amgen's commitment to safety and defending the current use and pricing of its drugs. Chief executive Kevin Sharer, for example, has asserted that many hundreds of thousands of patients who could benefit from Amgen's anaemia drugs are not receiving them — and pointed out that the price of Aranesp has fallen by one-third since it was introduced.

The company's travails have also served to remind the markets of the inherently high risks of the biotech sector, analysts say. "We're not talking about Pfizer, with 50 products in its portfolio," says Knight. "Biotechnology companies are vulnerable to disastrous events in one or a small number of their drugs." ■



Power games

Can motor racing go green? **Andreas Trabesinger** asked Max Mosley, head of Formula 1, how he wants the sport to develop energy-efficient technology that will also work in road cars.

When Robert Kubica moved to overtake a rival at the hairpin bend in the Canadian Grand Prix on 10 June, he lost control of his Formula 1 car and smashed head-on into a wall in a truly horrific-looking crash. His BMW-Sauber was travelling at 280 kilometres per hour as he tried to pass Jarno Trulli's Toyota, and after ricocheting off the barriers, the car somersaulted along the track before coming to rest with only one wheel still attached. Remarkably, Kubica emerged from the crushed shell of his car with mild concussion and a sprained ankle. His slight injuries are a testament to safety improvements in Formula 1 cars, and to the commitment made by the Fédération Internationale de l'Automobile (FIA), the sport's governing body, to safety standards. Kubica almost certainly would not have survived a similar crash 15 years ago.

The Montreal event, as with all 17 races held in this year's Formula 1 championship, is about the thrill of pushing automotive technology to the very edge of reason. Making sure that

the speed seekers are reined in and the sport stays within sensible limits is a difficult task in a contest of such extremes.

This task is the responsibility of the FIA, which until recently worried mostly about drivers' safety while keeping the race exciting enough to satisfy the tens of thousands of spectators at the circuit and the tens of millions of television viewers. But this heady mix of reason and adrenaline can have unexpected results. Last year, the FIA set out a 'green agenda' for Formula 1, announcing its intention to turn a sport in which cars guzzle 60 or 70 litres of petrol every 100 kilometres into a catalyst for greener technology for road cars.

Max Mosley is the man behind the wheel of the green agenda. In a penthouse high above London's Trafalgar Square, he lays out goals for the FIA to reach by 2009 and beyond. Now in his sixties, Mosley graduated from the University of Oxford, UK, with a physics degree, before going on to study law. He admits that

he is no expert when it comes to car technology, but he has been active in motor sports as a driver and team owner since the mid-1960s, and has been president of the FIA since the early 1990s. Mosley's vision of how Formula 1 will contribute to green technologies is simple:

SCHLEIGER/MILCH/CORBIS

make the research done in Formula 1 relevant to road cars, in particular reducing their emissions of carbon dioxide.

"Brilliantly clever, amazing engineering but utterly pointless, and irrelevant to the real world."

— Max Mosley

So how does Formula 1 plan to get there? The FIA has a powerful advantage in that it can rewrite the technical rules for the championship every year. In the past, the FIA restricted the power a car's engine was allowed to produce for safety reasons, typically by limiting the engine size. For the race engineers, the task was to extract the maximum possible power from a given size of engine (see 'Racing through the decades', overleaf), thereby ensuring that Formula 1 remains the fastest form of racing on a twisted circuit. But by the start of the 2011 season, Formula 1 teams will have to crack

a new technological nut: making the most of a given amount of energy. From then, the amount of fuel the cars can use in each race will also be restricted.

For Mosley the link with road cars is obvious: "This is precisely the problem that the car industry is trying to solve and indeed the world is trying to solve." He adds, "As soon as you look at it like that, you say 'why didn't we do this years ago?'" The reason, he says, is the same as why the road-car industry hasn't done it and that the public hasn't demanded it, because energy is still very cheap.

The links between Formula 1 and road cars have strengthened over the past decade. Of the 11 teams racing today, six are sponsored directly by major road-car manufacturers, only two of which — McLaren-Mercedes and Ferrari — were running their own teams in 1997, although many manufacturers were involved in the sport as suppliers of engines and other parts. The change came as the road-car industry embraced Formula 1 as a marketing platform, and its involvement has in turn benefited the sport as the costs of racing started to outstrip the available resources. Owning a Formula 1 team is a luxury few can afford, with running costs of up to hundreds of millions of dollars a year. Thirty years ago, the change of a single gearbox could require extra fundraising, but today the sport is flush with money from big-name sponsors and advertising.

Over-engineering

What has that money achieved? According to Mosley, until the FIA froze engine development at the end of last year's season, an average of 4 milliseconds of lap time were gained for every million dollars spent on engine development, and 20 milliseconds for every million dollars spent on optimizing the aerodynamics. Mosley is clear in his verdict: "Brilliantly clever, amazing engineering but utterly pointless, and irrelevant to the real world, because the engines were inherently inefficient." He points out that the teams have massive wind tunnels, supercomputers and model shops and they work 24 hours a day just to refine known technology. "This I want to stop," he says. "Let's get the really clever people working on the problem the whole world is trying to solve — which is just as good for Formula 1."

There are two areas in which Mosley thinks Formula 1 can make a lasting contribution to road-car technology, and that in turn will ensure the lasting success of the sport. These will be to recover energy lost through waste heat and braking. About two-thirds of the fuel energy in a car is lost as heat into the atmosphere — through exhaust gases and coolants. The other third propels the car forwards, but some of that kinetic energy is also lost, ultimately turned to heat, when the driver brakes. From 2009, new regulations for Formula 1 will allow, and thus force, the teams to recover a restricted amount of energy lost in braking, and use it to propel the car. The harder task of recovering the two-thirds of heat lost to the atmosphere is deferred until new regulations are introduced for 2011.

At present, the teams are not allowed to recover braking energy because of concerns about how the technology would perform under the extreme forces experienced by a Formula 1 car. The technology that does this is called a kinetic energy recovery system (KERS), better known to drivers of hybrid vehicles as 'regenerative braking'. In a modern hybrid car — which has both a petrol engine and an electric motor — the motor's batteries can be charged by either the petrol engine or regenerative braking. The energy can be stored in different forms, but the most viable options for Formula 1 seem to be electrical storage in batteries or capacitors or the use of a flywheel.

Although the 2009 regulations will not limit the cars' consumption of fuel directly — and refuelling will still be allowed during the race — the ability to regain kinetic energy means extra power for racing. In short, the car gains energy without having to carry extra fuel, and therefore weight. Another advantage of KERS is that the stored energy can be used to improve performance, especially during acceleration out of corners or overtaking of other drivers, giving racing fans a more exciting spectacle. Together, these factors make KERS extremely attractive to Formula 1 engineers.

Electric dreams

Burkhard Göschel, chairman of the FIA Manufacturers' Advisory Commission and former board member of BMW, is the 'technical brain' behind FIA's green agenda. He expects that most teams will go for electrical storage systems, either in the form of so-called supercapacitors (which have very high energy density and can store and release energy quickly) or new battery technology based on lithium-ion batteries.

Long term, both Mosley and Göschel are betting that the car industry will move towards using more electric power. "The electrification of the automobile can be anticipated, there is no way back. We are exactly on the right track with Formula 1, and road cars will follow this track," says Göschel. He is convinced that Formula 1 will make electrical energy-storage systems more efficient, smaller and lighter, and that the technologies

RACING THROUGH THE DECADES

Formula 1 in the 1950s

Fatalities: 8
Engine sizes: 1.5–4.5 litres
Power: 270 horsepower in 1958

Formula 1 in the 1970s

Fatalities: 10
Engine sizes: 1.5–3.0 litres
Power: 485 horsepower in 1974

Formula 1 in the 1990s

Fatalities: 2
Engine sizes: 3.0–3.5 litres
Power: 755 horsepower in 1997

FOX PHOTOS/GETTY IMAGES

ALLSPORT UK/GETTY IMAGES

M. BRANDT/BONGARTS/GETTY IMAGES

One formula for zero emissions

Can racing become emission free? A small Dutch company based in Amsterdam wants to create a race series, called Formula Zero, that will be based on cars powered by hydrogen fuel cells. Unlike conventional engines, fuel cells produce energy by reacting hydrogen fuel stored in a pressurized tank with oxygen taken from the air, so water is the only exhaust product.

It is still early days for Formula Zero. Founded in 2003 by Eelco Rietveld, an industrial design engineer, and Godert van Hardenbroek, an environmental consultant, the company is planning a race series for hydrogen-powered go-karts, which it hopes will kick off in 2008 or 2009. So far it has persuaded

several university teams to build fuel-cell go-karts, and it has earned an FIA-endorsed speed record for a fuel-cell vehicle that weighs less than 500 kilograms.

Last year, its go-kart, pictured here, reached an average speed of 61 kilometres per hour over 200 metres from a standing start.

But its long-term goal is more ambitious: a race series for car makers to showcase zero-emission technologies in full-sized race cars. Van Hardenbroek says they fully support the FIA's drive towards fuel economy: "Max Mosley made a



very wise move; the world will move towards hybrid cars, and Formula 1 should reflect that." But he recognizes that the technology has a long way to go to compete with petrol engines: "For

Formula 1 it would be very hard to make a transition towards fuel cells, this is not an incremental step." **A.T.**

developed on the way will be directly relevant to road cars. For example, the batteries used by hybrid fuel-electric vehicles are still too heavy, and the amount of energy that can be put in and taken out of a storage device is limited — problems that Formula 1 research, with its short design cycle and high-performance goals, seems ideally suited to fix. Mosley is confident that the race engineers will deliver: "You can't say you'll have it ready in two years, because the teams say they need it next week. The people in the next garage will have it next week."

So is Formula 1 heading in the same direction as the road-car industry? Paul Eisenstein, publisher of *TheCarConnection.com* in Pleasant Ridge, Michigan, and observer of the automotive industry since 1979, has no doubt that the car business is under enormous pressure to improve fuel economy; at the same time however, consumers are not willing to compromise on car size or performance. For these reasons, says Eisenstein, hybrids are not doing as well in practice as on paper: "The cost is high, and the performance of many models mediocre, in particular in terms of fuel economy. Most existing hybrids don't deliver what they promise; that's not good."

The next breakthrough for hybrid vehicles will have to come from making the interplay between the electric motor and the petrol engine more efficient, says Eisenstein. What will race engineers contribute to hybrid technology? "I see no reason why race-car technology shouldn't make it into road cars," says Eisenstein, "but such technology will have to meet tough criteria: What is it going to cost?

How long is it going to last? Nowadays, such components are expected to deliver at least 100,000 miles." Whether the technologies developed for Formula 1 will deliver both performance and durability, at reasonable cost, remains to be seen.

Heat treatment

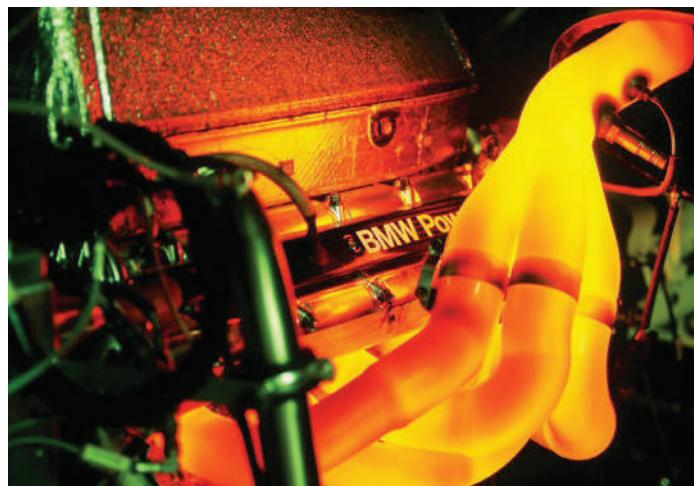
Hybrid vehicles would benefit from improved regenerative braking, but the recovery of kinetic energy is still playing with only a third of the energy contained in the fuel that the car burns. There is still the two-thirds lost as heat to think about. Getting that energy back is attractive, says Göschel, but not as simple to address. Formula 1 cars have previously harnessed 'turbocharging' technology to improve the engine efficiency. In a turbocharger, exhaust gases drive a turbine that compresses

the air flowing into the combustion chambers, and thus, eventually, allows fuel to be burned more efficiently.

Turbochargers were used by the teams during the 1980s, before being banned in 1989 because they gave engines dangerously too much power. The changes to the FIA regulations for 2011 onwards could provide a chance to bring the turbochargers back, but they have yet to be framed. What Formula 1 will bring to turbocharger technology for road cars — widely used in vehicles from turbodiesels to high-performance sports cars — is far from clear.

Charging up

More speculative are new ways to transform waste heat directly into electrical energy by use of physicochemical processes, but Göschel notes that such devices have very low



The heat is on: from 2011, Formula 1 teams will be able to reuse waste heat from the car's engine to boost performance.

BMW

efficiency. Further developed is a steam turbine that BMW introduced under the name of ‘turbosteamer’ — who would have thought that one of those could ever be discussed in the context of a Formula 1 car? — which is powered by the heat created by the petrol engine, so mechanical energy is recovered from heat. In a similar device, known as a ‘turbo-compound’, the exhaust gases drive not only the turbine of a turbocharger, but also a turbine in the stream of exhaust gases whose extra power can be used either directly or stored electrically. Unlike turbochargers, none of these devices is yet in production for road cars.

How have the Formula 1 engineers reacted to these rule changes? “The teams don’t like it, because we ask them to stop doing things they understand, and do things they don’t understand,” says Mosley. Göschel has noticed a more positive trend: “In the very beginning, our engineers had some concerns, but now there is a lot of excitement in working on new technology.” Nick Fry, head of the Honda Racing F1 Team, hopes that the rule changes will challenge young engineers, in particular, to come up with new solutions: “It’s an investment in people, in learning and in intellectual property. By pushing this type of technology where we have to perform publicly every two weeks, we must advance very quickly.”

Alternate take

But why stop with efficient energy recovery? Formula 1 could switch to using biofuels, maybe starting in 2011, says Mosley: “We would like to use a biofuel. The question is, which one. There are so many competing biofuel systems.” What Formula 1 might end up doing is taking whatever fuel becomes adopted more widely, rather than picking a fuel in advance. Fuel cells relying on hydrogen are not yet being considered for Formula 1,



Many hands make light work: short design cycles put Formula 1 engineers under pressure to deliver.

although a small Dutch company is trying to launch a fuel-cell race series (see ‘One formula for zero emissions’). In addition, the FIA has an Alternative Energies Commission that organizes an annual cup race with vehicles that use alternative energies.

Mosley is planning to step down as FIA president at the end of his fourth consecutive term in October 2009, so is this green agenda all about his legacy? He admits that it plays a part, but he

compares today’s environmental concerns (See ‘Carbon credentials’) with the safety con-

“We are exactly on the right track with Formula 1, and road cars will follow.”

— Burkhard Göschel

cerns that dominated Formula 1 when he first became president of the FIA. “It’s a little bit like the safety debate, in that you work on safety because you don’t want to kill anybody, you don’t want anybody to get hurt, but also, society won’t permit you to kill people like we did in the 1960s.” During that period a driver died every year in Formula 1. “So, you’ve got two reasons: you want to do it yourself, but also you have to have regard to what society allows you to do.”

Will Formula 1 be perceived as a ‘green sport’ in the future? “I don’t know whether the fans will like it,” says Mosley but he doesn’t think that reason and adrenaline are incompatible. As people become increasingly conscious about carbon emissions and fuel economy, he hopes they will still be fascinated by a very fast, very powerful — but fuel efficient — Formula 1. In general, Mosley is pragmatic about the effect of the rule changes: “If it’s technically interesting, that’s fun, and if it makes a contribution to society, that’s good,” but ultimately he thinks Formula 1 needs public support in order to survive. “The number one thing is to make it so attractive and interesting that the public continues to pay for it.” ■

Andreas Trabesinger is an associate editor for Nature Physics.

Carbon credentials

Since 1997, the Fédération Internationale de l’Automobile (FIA), motor racing’s governing body, has supported a research project aimed at offsetting the carbon-dioxide emissions caused by Formula 1 teams (from the race cars themselves and from transporting teams to events) during a Grand Prix season. Through the FIA Foundation, a UK-registered charity, the FIA offsets annual emissions of the 11 teams racing — estimated in 1997

to be around 20,000 tonnes of CO₂ — by supporting the ‘Scole Te’ project, which helps communities in southern Mexico to develop sustainable land management and better livelihoods.

As of December 2005, 888 farmers from 43 communities across the states of Chiapas and Oaxaca were included in the project, says Richard Tipper, president of the Edinburgh Centre for Carbon Management, UK, which consults on the project.

Unlike other sporting events, such as the 2006 World Cup in Germany, the FIA does not offset emissions caused by fans who travel to the events, so it can’t claim to be carbon neutral. David Ward, director general of the FIA Foundation, says that the project’s effectiveness will be reviewed this year, and the foundation will review the carbon footprint for the Formula 1 teams to see whether it has changed since 1997. As yet, The FIA Foundation has no plans to go carbon neutral. A.T.

EXTREME CULTURE

From acid mine drainage to the bowels of the Earth, **Josie Glausiusz** reports how researchers

are taking great pains to grow recalcitrant bacteria.

To reach one experimental site — inside Lechuguilla Cave, New Mexico — Diana Northup and her team must venture 350 metres down into the Earth, hiking along several kilometres of obstacle-strewn passages. For days they rappel down pits, traverse narrow ledges bordering steep drops and pristine lakes, and clamber over boulders — all the while carrying heavy packs of equipment. When they finally arrive at their sampling areas, they must change into clean suits, including bonnets and sterile gloves. “It’s hard to convey to you how filthy you can become in four to five days,” Northup says. “You’re just constantly drenched in sweat. These crusts we study get all over you, gypsum gets all over you, and the sweat helps plaster it all onto you. It’s just lovely.”

Northup, a microbiologist at the University of New Mexico in Albuquerque, and the inter-university SLIME team (Subsurface Life in Mineral Environments) are looking to find how cave bacteria deposit oxidized iron and manganese crusts on the walls of underground caverns. Her first attempts to grow these microbes in the lab were thwarted. All that appeared on her agar plates were ‘weeds’, fungi that hikers had tramped into the caves on their boots. Hence the clean suits, to prevent contamination. To get the right incubating conditions, Northup came up with a simple solution: cultivate the bacteria in glass tubes

inside the caves themselves, providing the precise environment they prefer: total darkness, low temperatures and high humidity.

Northup’s efforts may seem extreme, but they illustrate the lengths to which some microbiologists will go to culture the seemingly unculturable. According to a common estimate, some 99.9% of microbes will not reproduce in Petri dishes lined with nutrient-rich agar — a culturing technique virtually unchanged since its invention in the early 1880s. Most have more stringent criteria, including rare minerals, specific biochemical signals or the organisms with which they usually cooperate. In other words, they need the comforts of home.

Difficulties in culturing bacteria have helped spur metagenomics, in which researchers descend on a sample — a bucket of sea water, a patch of human skin or a handful of soil — and sequence all the microbial DNA within. Such techniques can start to reveal how many bacterial species can be found in any spot, or how many gene variants appear in the collective population. For example, the J. Craig Venter Institute announced in March 2007 that its *Sorcerer II* Global Ocean Sampling Expedition had discovered 6 million new genes and thousands of new protein families. Mitchell Sogin’s group

at the Josephine Bay Paul Center of the Marine Biological Laboratory in Woods Hole, Massachusetts, has been sequencing 16S ribosomal RNA from seawater samples to measure species numbers and found microbial diversity much higher than had been reported before.

Exciting as these results are, they also highlight the culture gap. Some biologists remain dedicated to taming recalcitrant bacteria in the lab, challenged and sometimes aided by

“They’re not uncultivable; they’re just not cultured yet.”

— Lynn Rothschild

metagenomic findings. Using specially designed diffusion chambers, mud-filled vessels, microarrays and bioreactors, they mimic the conditions in which the bacteria live naturally. They are growing what no one has grown before, which should reveal more about the organisms than a mere gene census, says evolutionary biologist Lynn Rothschild of NASA’s Ames Research Center in Moffett Field, California. “Just because they have a gene, they may not use it. Or we may not learn when they use it,” she says. “Unculturability is not a biological characteristic, it’s a human failing. They’re not uncultivable; they’re just not cultured yet.”

Some researchers have what Rothschild calls “a green thumb for microbes”. Belinda Ferrari of Macquarie University in Sydney, Australia, has developed a novel microcultivation

K. INGHAM

method, using a mud slurry as a medium for growing soil bacteria. Using this system, she has cultivated a wide diversity of bacteria: they include those called TM7, found in soil, bio-reactor sludge and the human mouth. Nutrients come only from the mud, a source ideally suited for these microbes, which may suffer in the high sugar levels that suit lab 'weeds'. "If you're in a pot full of jam and not much else, you'll just overdose and get sick from too much sugar," Ferrari explains.

From culture to cash

There are commercial spinoffs for extreme culture. In the Netherlands, microbiologists Mike Jetten and Marc Strous of Radboud University Nijmegen have enriched anaerobic ammonium-oxidizing (anammox) bacteria and nitrate-dependent anaerobic methane-oxidizing bacteria. The latter, collected from a canal contaminated with agricultural runoff, were grown in a bioreactor with a nutrient-limited supply of carbon dioxide, nitrate and methane. The two succeeded in culturing colonies of these microbes after a year of trial and error. Anammox bacteria are now being used to clean waste water on a large scale in the Netherlands and Japan.

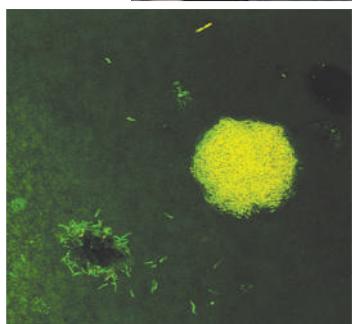
Meanwhile, Kim Lewis and Slava Epstein of Northeastern University in Boston, Massachusetts, are trolling the depths of the uncultured for new antibiotics using a diffusion chamber in which microbes are suffused in the conditions of their natural environment — soil, for example, or sea water plus marine sediments. Lewis and Epstein have founded the company NovoBiotic to capitalize on their cultures. "The practical benefits are enormous," Lewis says. "If you want to discover new stuff, you want to go to organisms you haven't seen before. It's reasonable to assume that 99% of the remaining bacteria will have at least some useful antibiotics."

The spelunking Northup has begun to find antibiotics in her samples. Initially, though, she strove to grow crust-forming cave bacteria to understand their basic biology. "One of the reasons we culture rather than do DNA sequences is because we want to catch them in the act of precipitating the minerals, so that we can say definitely, 'These guys can do it,'" she says.

While underground she and her team scrape crust off the cave wall and stab it into a glass tube filled with 'sloppy' agar. An iron carpet tack or reduced manganese at the bottom of the tube provides metal, and rock dust from limestone in the caves provides trace minerals that the microbes need. The bacteria form



SLIME's lab in a cave (above);
Ferrari's colonies of bacteria
around soil particles (left).



bands in the agar at different oxygen levels and precipitate iron and manganese deposits.

Because these microbes grow so slowly, Northup and her colleagues usually leave their cultures percolating for years. She has already cultivated a variety of microbes from New Mexico caves, including iron- and manganese-oxidizing species of *Bacillus*, *Caulobacter* and *Alcaligenes*: odd-looking bacteria whose strange morphology she refers to as "beads on a string" and "hairy sausages". She has also found samples of *Actinomycetes*, bacteria that are known to produce more than 4,000 antibiotics, as well as enzymes that could be of use in biotechnology.

Antibiotics and ecology

Northup now has a student searching her *Actinomycetes* samples for novel antibiotics. "Because this is fairly unexplored habitat, we're hoping to find some rarer ones," she says. "We'd also maybe like to characterize what habitats in caves tend to be richer in organisms that produce antibiotics — to test the hypothesis that in low-nutrient environments they're more likely to produce these secondary metabolites to keep their neighbours at bay."

Such ecological questions could have widespread implications. A group led by Steve Giovannoni, a microbiologist at Oregon State University in Corvallis, has cultivated at least

11 strains of SAR 11, a marine microbe that is not only the smallest independently replicating bacterium known but also has the smallest genome ever seen in a free-living cell. What's more, SAR 11 "is possibly the most abundant organism in the oceans", says Giovannoni. It converts organic matter into carbon dioxide, and therefore plays a huge role in global biogeochemical cycles. All the more reason, therefore, to understand its basic biology — something Giovannoni thinks is best done through pure culture. "Metagenomics is fantastic," he says, but "with an organism like this, it's vastly superior to have a culture. If hypotheses emerge from a genome sequence, you can then test them. Also, at least 30% of the genes in every microbial genome cannot be identified under the best of circumstances."

Even as some microbiologists seem aligned with pure culture techniques, others have used metagenomics as a guide in raising obstinate microbes.

Jill Banfield of the University of California, Berkeley, and her colleagues

used 'shotgun' sequencing methods to characterize five species of iron-oxidizing bacteria and archaea in biofilms from an acid mine-drainage system in Iron Mountain, California, and discovered that only one carried the genes that code for enzymes to fix nitrogen. "We got a pretty complete inventory of the pathways in each organism that's present," Banfield says. "Because we had near-complete genomes we could say, 'This organism — ha ha! — has the ability to also fix nitrogen, whereas the others don't.'" By raising them in a sulphuric-acid bath supplemented with iron, with no nitrogen except the gaseous form, the team was able to isolate the nitrogen-fixing bacterium, *Leptospirillum ferrodiazotrophum*.

As metagenomic efforts ramp up, more of these extreme-culture techniques are likely to follow because the two approaches inform each other. "You get totally different bits of information," says microbial ecologist Anna-Louise Reysenbach of Portland State University in Oregon, who for the first time has cultured an acid-loving, thermophilic, sulphur-reducing species of archaea called *Aciduliprofundum boonei* from deep-sea vents at Valu Fa Ridge between Tonga and Fiji. "I love being able to have the organism in culture, but also knowing what its distribution is like. I definitely think that both have great value. There shouldn't be one without the other."

Josie Glasiusz is a science journalist and author based in New York.

K. INGHAM

B. FERRARI, N. TUJULA

Scientists should promote co-operation, not boycott

SIR — Since 1931, the International Council for Science (ICSU) has upheld the principle of the universality of science, based on the right of scientists to work without discrimination on the grounds of citizenship, religion, creed, political stance, ethnic origin, race, colour, age or gender.

The entire ICSU membership, representing the scientific community in 112 countries and all disciplines, has consistently expressed its unequivocal support for this principle. This stance has stood the test of time throughout the Cold War, apartheid in South Africa and the new challenges posed by international terrorism. It is a strong expression of solidarity across the international science community: a critical reference point for individual scholars confronted with threats to their freedom.

The decision by the congress of the UK University and College Union to recommend that its members bar academic exchanges with Israeli researchers is a flagrant breach of this principle. It has rightly drawn substantial adverse comment from scientists, newspaper columnists and human-rights activists in the United Kingdom and internationally.

It is easy to understand the strong feelings generated by conflicts and people's desire to demonstrate their opposition to the actions of governments. But to do so through the medium of individual scholars is to sacrifice a profoundly important principle of freedom and solidarity. In situations of strife and conflict, it is surely the duty of scientists to promote international understanding and co-operation — not to penalize each other for the shortcomings of their governments.

Bengt Gustafsson

ICSU Committee on Freedom and Responsibility in the Conduct of Science, and Department of Astronomy and Space Physics, Uppsala University, Box 515, SE-751 20 Uppsala, Sweden

US government enforces boycott of whole countries

SIR — The US Educational Commission for Foreign Medical Graduates (ECFMG) processes the applications of foreign students for the US Medical Licensing Examination, which evaluates candidates' basic and clinical knowledge in medicine. The examination has two steps, both of which must be passed for the ECFMG to recognize the individual as a medical doctor in the United States.

As a medical student, I applied for step 1 in January 2007, entering my country (Iran) in the contact address section. A message in red letters appeared on the registration page: "It

has come to our attention that ECFMG may be subject to specific United States federal regulations that prohibit entities from doing business with or providing a service to any individuals who have an address of residence in specific restricted countries. The country of Iran is included on this list of restricted countries. In light of this, ECFMG is not able to allow you to request this service."

The political status of Iran or any other nation is not relevant to education. The result of the restriction is that all medical students and graduates who live in Iran are prevented from taking the US licensing examination. This surely is academic discrimination against a whole country.

Sina Zarrintan

Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

Cincinnati's rhino breeders bring home the bacon

SIR — Recent welcome successes in Sumatran Rhino breeding at the zoo in Cincinnati, Ohio (*Nature* 447, 125; 2007), prompt the idea that Cincinnati's existing colloquial name, 'Porkopolis' (reflecting the city's historic importance in the pig trade), could be replaced by 'Rhinopolis'.

Martin F. Heyworth

Veterans Affairs Medical Center, University and Woodland Avenues, Philadelphia, and Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Cooling may be possible, but we need safety data

SIR — The global cooling geo-engineering ideas discussed in your interesting News Feature 'Is this what it takes to save the world?' (*Nature* 447, 132–136; 2007) merit, in my view, further study to determine their operational viability and associated meteorological and other ramifications.

I proposed such a global cooling scheme some years ago (*Nature* 347, 339–340; 1990), involving inadvertent albedo enhancement of shallow maritime stratiform water clouds, which cover about one-sixth of the global surface, by spraying seawater particles of diameter about 1 micrometre from close to the ocean surface, underneath such clouds. A significant fraction of these would be transported by turbulence into the clouds, where they would act as efficient cloud condensation nuclei, creating additional droplets and thereby increasing cloud reflectivity.

Since then, this suggestion has been investigated further (see K. T. Bower *et al.*

Atmos. Res. 82, 328–336; 2006, and references therein). One advantage of the scheme is that it constitutes a controlled version of a process that occurs naturally, in that seawater particles are continually being produced at the ocean surface by wave-breaking and bubble-bursting, and some fraction of these rise into the stratiform clouds above and produce additional cloud droplets. Other advantages are that it can be switched off immediately, with conditions returning to normal within a few days; all required energy is derived from the wind; and the only raw material is seawater.

There is no justification for deploying this or any other global cooling scheme, however, until rigorous examination of all possibly deleterious meteorological and other ramifications have been conducted and found acceptable.

John Latham

Mesoscale and Microscale Meteorology Division, National Center for Atmospheric Research, PO Box 3000, Colorado 80307-3000, USA

Darwin Centre will be fit for its range of purposes

SIR — Your News story 'Anger at "unfit" museum design' (*Nature* 447, 239; 2007) reports some people's belief that the design of the Natural History Museum's Darwin Centre Phase Two is unfit for purpose, as the building will not have room to house the entire insect and plant collection.

The crux of the issue is that the public is being given access to our science, and this takes space. Building new facilities for the entirety of our collections, research and public access in one go is not feasible, with current funding. Instead we are taking it by stages. With the completion of Darwin Centre Phase Two, we will have more than half of our 70 million specimens in high-quality storage.

Balancing the needs of collections, research and public access will help us to advance knowledge of the natural world and to communicate this to the public. It will mean some changes in how we work and it will take time to adjust. However, we think that the new building will lead to a long-term improvement in the way we carry out and communicate the museum's work in taxonomy and systematics.

We have received a high level of support for this project, and are confident that it will be a real benefit to the museum, the scientific community and the wider public.

Richard Lane

Natural History Museum, Cromwell Road, London SW7 5BD, UK

Contributions to Correspondence may be submitted to correspondence@nature.com.

BOOKS & ARTS

Weird and wonderful

The quantity and diversity of deep-sea life are a match for the rainforests.

The Deep: The Extraordinary Creatures of the Abyss

by Claire Nouvian

University of Chicago Press/Fayard: 2007.
256 pp. \$45/€40

The Silent Deep: The Discovery, Ecology and Conservation of the Deep Sea

by Tony Koslow

University of Chicago Press/University of New South Wales Press: 2007. 312 pp.
\$35/Aus\$49.95

Mark Schrophe

The 'vampire squid from hell', the fireworks jellyfishes and the pigbutt worm are just a few of the creatures of the deep sea that have remained unseen by all but a select few. Two new books offer complementary views of this strange expanse and its inhabitants.

No photo collection could replicate a visit to their realm or the breadth of the diversity to be found there, but Claire Nouvian's *The Deep*, with more than 200 large-format photos, comes closer than any previous book. *The Silent Deep*, by deep-sea biologist Tony Koslow, is an excellent companion, with textbook depth on all aspects of deep-sea science and conservation.

Koslow begins this exhaustively researched and referenced volume with a historical review stretching back to Socrates, who posited that the deep was a lifeless desert where "nothing is in the least worthy to be judged beautiful by our standards". If only he had known.

The Socratic view was shattered by the UK *Challenger* expedition, which began in 1872 and laid the foundations of modern oceanography. More than a few readers will long for the days when, as happened with that famed voyage, an appeal to national pride — in the pages of *Nature*, no less — and a one-page scientific plan was enough to spur sufficient funding for a multi-year global expedition.

Koslow proceeds to what is now known and, just as significantly, not known about deep-sea habitats and fauna, covering everything from abyssal plains to hydrothermal vents, and the seminal findings related to each.

The deep's limited food resources — faeces, dead phytoplankton, crustacean moults, discarded mucus and noxious chemicals — and its monotonous environmental conditions may not seem a recipe for diversity. But paradoxically, that diversity is vast, and includes numerous fisheries, as many stony corals as on



The deep sea's beautiful and bizarre creatures, such as this piglet squid, remain unseen by all but a few.

shallow reefs, and abyssal plains with as many species as tropical rainforests.

"The communities of the deep-sea benthos [the sea bottom]," Koslow writes, "pose the greatest challenge of any ecosystem on Earth to our understanding of the planet's biodiversity." He explains that the paradox is probably created by a variety of factors, such as low productivity, which prevents any single species from becoming numerous enough to drive out its competitors.

The book's final section covers issues of human exploitation of the deep, potential impacts, and Koslow's view of what might be done to minimize those impacts. Although his passion for and fascination with the deep sea shine through, his tone is one of reasoned consideration of the data, rather than that of a scientist turned activist. This lends weight to his arguments, including a call for expanded international regulation of deep trawling on the high seas.

The Silent Deep should be accessible to a broad audience, but the author does assume readers will have an understanding of basic

scientific concepts. In contrast, *The Deep*, with its visual focus, targets everyone and seems bound to accomplish one of Nouvian's stated goals: to reveal the creatures of the deep and interest the public in them so that they will be more inclined to protect them.

Nouvian became enamoured of the deep during a visit to the Monterey Bay Aquarium's pioneering deep-sea exhibit. This inspired a quest that would involve collecting more than 6,000 photos and diving to the bottom of the Gulf of Maine in a submersible.

The Deep offers spectacular views of such marvels as bioluminescent dragonfishes and a googly-eyed glass squid. Other animals have less colourful names, but are exquisitely beautiful. It is difficult to imagine anyone who would not be enchanted by the creatures on display.

Each organism receives proper treatment in Nouvian's clever extended captions. She describes, for instance, the naked sea butterfly's "somewhat barbaric table manners", which involve pulling sea snails from their shells with hooks and swallowing them whole, or the sea spider with a penchant for driving its proboscis

D. WROBEL

into prey "like a straw into a milkshake". The book also includes a collection of essays by prominent deep-sea biologists from around the world, on topics from trenches and seamounts to the varied uses of bioluminescence. These lend valuable context to the photos.

It is remarkable that almost all the photos are drawn straight from researchers, not professional photographers — the deep's creatures are beautiful, bizarre and at times even grotesque enough to capture attention without any fashion photography. This wealth of

photos just needed to be made available; in doing so, Nouvian has done both the field and the public a service.

Collectively, these books offer a spectacular visual and cerebral introduction to the wonders of the abyss that could awaken many to the idea that, as Koslow puts it, exploration and protection of the deep sea "is one of the great scientific voyages of discovery, one that humankind has only just embarked upon." ■

Mark Schrophe is a Florida-based freelance writer specializing in ocean topics.

in 1865 displayed the first gorilla specimens seen in Australia. McCoy wanted to disprove Charles Darwin's evolutionary theory by showing how different gorillas were to humans, and thought that installing an ape case would convince visitors that they could not be descended from apes. Crowds came to see it, making it one of the museum's most popular exhibits, even though Darwinism triumphed.

Another essay considers species that were known only from specimens in museum collections and were thought to have become extinct, such as Leadbeater's possum. Museum curators spent many years searching for it before it was finally rediscovered in 1961. The possums are now known to be relatively abundant in small areas of mountain ash forest, where they live high in the trees.

Australia's unique fauna continues to influence how people use the continent. Earlier this year, an iron-ore mining project in western Australia worth £5 billion (US\$10 billion) was blocked by government officials after several new species of spider-like troglobites were discovered at the site. These blind cave-dwelling species have evolved to live in the dark and are killed by even brief exposure to sunlight. They are the tiniest animals ever to have halted major construction work. ■

Colin Martin is a London-based Australian writer.

Wonders from down under

Continent of Curiosities: A Journey Through Australian Natural History

by Danielle Clode

Cambridge University Press: 2006.
224 pp. £35

Colin Martin

Male bowerbirds attract females by building bowers and decorating them with disparate, often highly coloured objects. Danielle Clode's book on Australian natural history, written after a stint as scientific interpreter at Museum Victoria in Melbourne, takes a similar approach to attracting readers. She has teased out stories from the museum, founded in 1856 and one of Australia's oldest, by exploring its collections and talking to curators. As the historian Tom Griffiths observes in his foreword, Clode's book reinforces "the essential link between research and collections in a period when museums increasingly have come to privilege the manager and designer over the research curator".

Years ago, I was enchanted to learn that in the late 1860s the poet Dante Gabriel Rossetti had kept a wombat in a menagerie at his London house, which amazed his cronies in the Pre-Raphaelite Brotherhood. Clode, on the other hand, looks at how Australia's unique wildlife influenced the development of its indigenous natural history, rather than how weird and wonderful antipodean fauna were received in the Northern Hemisphere.

Although Clode's bowerbird approach, triggered by the objects that caught her eye at Museum Victoria, is informative, her book is essentially an anthology of essays, and lacks a synthetic view. It was surprising that she included an essay on hominid fossils, none of which originated in Australia; and her riff on theories about the extraterrestrial origin of life hangs tenuously on the hook of a meteorite that hit near the Victorian town of Murchison in 1969. Her accounts of the theories of continental drift and biogeography are much more apposite.

The history of Melbourne's water and sewerage systems seems a quirky choice, but it is an interesting story and pertinent, given the current drought in Australia. From the early

1950s, Melbourne's water was sourced from extensive forest catchments, which purified it naturally. Later, museum researchers were important in monitoring the effects of dams on rivers. It had been thought that some types of dam might be better than others for improving invertebrate biodiversity downstream. It was shown, however, that the proximity of tributaries below a dam is a more important factor in invertebrate recolonization and keeping rivers healthy.

Humans selected by Clode include the museum's first director, Frederick McCoy, who



The first gorillas to be seen in Australia were displayed in 1865, exciting debate about human origins.

DNA cycle

Passengers travelling from London to Cambridge by train may glimpse this striped cycle path cutting through the flat countryside as they near their destination. Over a mile of its length, different coloured bands spell out the base-pair sequence of the *BRCA2* gene — mutations in which are associated with increased risk of breast cancer.

The path includes the UK cycle network's ten thousandth mile. John Sulston, director of the Cambridge Sanger Centre during its work on the Human Genome Project, and a keen cyclist, helped develop the idea of decorating it with a gene.



BRCA2 was picked because the gene contains about 10,000 base pairs, echoing the cycle network's celebratory theme. But there were other reasons, too. As a predictor of disease and the subject of a patent dispute, *BRCA2* highlights some of the social issues that surround gene sequencing.

"I expect this will be one of only a few human genes to be represented as a cycle path," says Michael Stratton, who identified and mapped *BRCA2* twelve years ago and acted as a scientific advisor to this project. It's a safe bet — were the entire human genome laid down at the same scale, the path would circle Earth about ten times. J.H.

A golden circle in the sky

An Ocean of Air : A Natural History of the Atmosphere

by Gabrielle Walker

Bloomsbury/Harcourt: 2007. 321 pp.
£15.99/\$25.

Raymond T. Pierrehumbert

For its passengers stopping over in Reykjavik, IcelandAir offers a one-day Golden Circle Tour of the principal sights — an hour at Thingvellir, a photo-stop at Dettifoss, enough time at Geysir to at least see Strokkur let loose, if not the old girl herself, and so forth. It's not a great way to see Iceland, but it gives travellers some impression of the wonders of the place. Gabrielle Walker's grand tour of Earth's atmosphere is very similar, although without the obligatory stop at the souvenir shop.

An Ocean of Air is at once better and worse than IcelandAir's tour. First, the good: Walker has a fine facility with words, and paints a thrilling picture of the grand sweep of atmospheric science and the parade of great minds and key experiments that led us to where we are today. Although there is little here that is not already accessible in other popular science writing, such as Lucy Jago's magnificent *The Northern Lights* (Hamish Hamilton, 2001), the way Walker assembles the collage is an achievement. Even if you know the territory well, it adds a lot to see the whole story laid out compactly with so much enriching historical and biographical detail.

However, although the Golden Circle tour guides may not tell you much, what they do tell you is generally true. The same, alas, cannot be said for Walker. I became suspicious in the first chapter, dealing with Galileo's determination of the weight of air. Walker starts well enough, discussing Galileo's pathbreaking experiments in weighing compressed air in a bottle, but

then declares that he believed that air became weightless when returned to its proper place, the sky, making him out to be some kind of neo-Aristotelian.

Could the father to the Renaissance rebirth of hydraulics really have been so misguided? Well, no, it turns out. In the section of the *Dialogo* dealing with these experiments, it is clear that Galileo is talking about Archimedes' principle. The flaw in his reasoning is more subtle and interesting, arising from Galileo's ignorance of what we would now call Newton's Third Law, of reciprocal actions. A bit of air in the atmosphere is made neutrally buoyant because the bit of air below pushes up on it so as to cancel its weight. But that means the air being held up pushes back on the air below, which must in turn push on what's below it, and so forth, until finally it can exert a force on a solid surface, or on water being pushed through a siphon.

This may seem an academic point, but one reads a popular science book not just for entertainment, but for illumination. If it leaves the reader with misconceptions, one might as well be reading science fiction instead. And there are lots of misconceptions here. Walker tells us that breathing pure oxygen would be fatal. It's not, and astronauts used to do it all the time; you can breath pure oxygen more or less indefinitely without harm, as long as the pressure is low enough.

She also tells us that Fourier computed what Earth's temperature would be without an atmosphere, and found it to be far colder than it actually is. Fourier did no such thing, and could not have with the tools of his time. By repeating this urban legend from secondary sources rather than delving into Fourier's writings, she misses the story of his remarkable, genuine contribution to our understanding of Earth's energy

balance (see *Nature* 432, 677; 2004).

Walker says that Gilbert Plass restored concern about CO₂ and global warming by showing that its absorption bands are saturated (absorbing as much infrared as they can, so that addition of more CO₂ causes no more absorption) only at sea level, not at upper atmospheric pressures. There is indeed a story to be told about the 50-year struggle to understand how radiation travels through the atmosphere, but this isn't it. Carbon dioxide absorption is not saturated at sea level, and would not be even if the CO₂ content were 10,000 times greater than today. The real problem with the prevailing conception of the greenhouse effect had more to do with a failure to appreciate the role of cold air aloft, and little to do with either pressure broadening or band saturation.

And when it comes to long-term CO₂ changes, Walker spins a fanciful yarn around a highly speculative theory involving dinosaurs' supposed ability to digest lignin in wood (they almost certainly couldn't), rather than discussing the well-established role of the formation of limestone by silicate weathering. There is so much sloppiness of this sort that when, at the end of the book, Walker informs us that the Inuit of Greenland are unconcerned about global warming because they are hunters, not farmers, we hardly flinch.

She stops short of some of the exciting current research — notably the role of the exosphere in hydrogen escape from the atmosphere, which has profound implications for the origins of life — but the main problem is what she gets wrong, not what she leaves out. The too often lackadaisical research is a real pity. Read this book for the grand sweep of history, but keep your Web browser and library card handy to sort out which bits to believe. ■

Raymond T. Pierrehumbert is Louis Block Professor in the Geophysical Sciences, the University of Chicago, 5734 South Ellis Avenue, Chicago, Illinois 60637, USA.

EUKARYOTE EVOLUTION

Engulfed by speculation

The notion that eukaryotes evolved via a merger of cells from the other two domains — archaea and bacteria — overlooks known processes.

Anthony Poole and David Penny

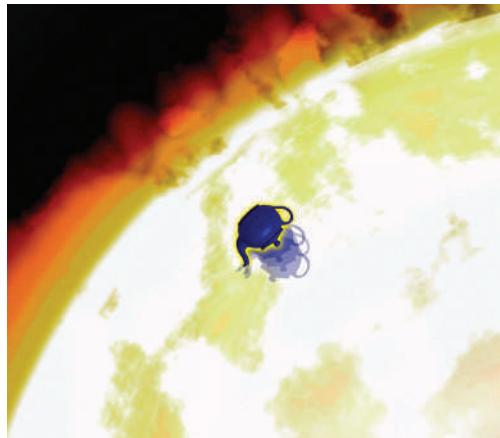
In the absence of direct evidence, science should proceed cautiously with conjecture. Geologist Charles Lyell (1797–1875) warned us not to proceed like medieval scholars, who “often preferred absurd and extravagant positions, because greater skill was required to maintain them”. Scientific speculation, Lyell emphasized, must take known processes into account. This has not happened with the debate on how eukaryotes (animals, plants, fungi, protists) arose. The conflicting hypotheses currently on offer show a curious disregard for mechanism.

One thing at least is agreed: the mitochondrion, powerhouse of the eukaryote cell, evolved from an engulfed bacterium. The question is ‘who’ did the engulfing. Did an archaeon engulf a bacterium? Did a bacterium, bacterial consortium, or RNA cell engulf first an archaeon (which became the nucleus) and then the mitochondrial ancestor? Perhaps nuclei emerged in a virus-infected archaeon, which then engulfed mitochondria. Which, if any of these, is right?

In the mid-1990s, a somewhat pedestrian view of eukaryotic origins, the ‘archezoa hypothesis’, held sway. This maintained that a protoeukaryote (with nucleus) engulfed the mitochondrial ancestor. Supporting the theory were ‘archezoa’, anaerobic eukaryotes with no mitochondria. Archezoa apparently populated the oldest branches of the eukaryote tree, suggesting that eukaryotes began diversifying before mitochondria entered the picture.

The archezoa hypothesis is thus composed of two independent hypotheses: (a) that a protoeukaryote host (PEH) engulfed the mitochondrial ancestor, and (b) that modern archezoa are ‘missing links’ that never possessed mitochondria. Hypothesis (b) is now unanimously rejected: every archezoan examined bears vestigial mitochondria, or genes inherited from mitochondria. Thus, all modern eukaryotes evolved from a mitochondrion-bearing ancestor.

But the baby was thrown out with the bath-water. Hypothesis (a) was also rejected, and because eukaryotes and archaea share a number of similar genes, the deposed PEH was replaced with archaea. Consequently, incorporation of the mitochondrion — not the origin of the nucleus — was hailed as



the defining event in eukaryotic origins. This opened the floodgates of speculation, and numerous new hypotheses emerged. None is supported by observation: no archaea reside within bacteria, no bacteria reside within archaea, viruses have preposterously few similarities to the nucleus, and no RNA cells exist.

Of course, missing links might exist that could bolster one of the new theories. Because it is not possible to examine every cell on the planet for evidence, proponents could always argue that their theories are not disproved. Should we take them seriously, then?

No. Recall Bertrand Russell’s (1872–1970) metaphorical teapot, orbiting the Sun but too small to be detected by telescopes. Without evidence, we cannot say the teapot is *not* there. Russell quipped, “if I were to go on to say that, since my assertion cannot be disproved, it is intolerable presumption on the part of human reason to doubt it, I should rightly be thought to be talking nonsense”. In other words, the onus is on proponents, not sceptics, to find evidence for their theories.

If archezoa are not missing links, has evidence for the PEH theory also vanished? Are the various theories on equal ground? Again, no. First, phagocytosis — one cell engulfing another — is widespread among eukaryotes, but unknown in bacteria and archaea. Eukaryotes engulf bacteria as food, and some engulfed bacteria can evade digestion. There are conditional endosymbioses, in which one cell lives within another but escapes under adverse conditions, and stable endosymbioses, where host and guest are locked into the same evolutionary trajectory.

Predatory eukaryotes even pilfer chloroplasts from their hosts. Second, every other organelle that evolved from a once free-living cell was initially engulfed by a eukaryote. This is, after all, how the chloroplast evolved.

The origin of mitochondria under the PEH theory is not some freak event requiring special explanation; mitochondria simply descend from an engulfed cell co-opted as an organelle. That this was an evolutionary success is clear: bona fide archezoa no longer exist. In contrast, archaea are not known to engulf bacteria (nor vice versa). So the alternatives to the PEH hypothesis require that bacterial or archaeal ancestors ‘invented’ engulfment, only to lose it again without trace — another assertion that cannot be disproved.

As Norman Pace pointed out (*Nature* 441, 289; 2006), eukaryotes did not evolve from archaea; they share a common ancestor. This fits the PEH theory — the host was a direct ancestor of modern eukaryotes. If the host were, say, an archaeon, eukaryotes would appear on the tree of life as a subdivision of archaea — they would be phylogenetically part archaeon, part bacterium. The tree tells us they are part eukaryote, part bacterium, and sister to archaea.

Much uncertainty surrounds eukaryote origins. How did phagocytosis and the nucleus evolve? How was endosymbiosis between mitochondrion and host established? In answering these questions, we need the PEH theory. It is the only explanation based on a host capable of engulfing the mitochondrial ancestor by known processes — rather than by mechanisms founded in unfettered imagination.

Anthony Poole is in the Department of Molecular Biology and Functional Genomics, Stockholm University, SE 106 91 Stockholm, Sweden. David Penny is at the Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular Biosciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand.

FURTHER READING
 Cavalier-Smith, T. *Int. J. Syst. Evol. Microbiol.* **52**, 297–354 (2002).
 de Duve, C. *Nature Rev. Genet.* **8**, 395–403 (2007).
 Poole, A. M. & Penny, D. *BioEssays* **29**, 74–84 (2007).
 Van der Giezen, M. & Tovar, J. *EMBO Rep.* **6**, 525–530 (2005).

CONCEPTS

NEWS & VIEWS



R. WILKING/REUTERS

An ethanol plant in Colorado, surrounded by fields of corn (maize).

CHEMICAL ENGINEERING

Hybrid routes to biofuels

Lanny D. Schmidt and Paul J. Dauenhauer

Traditional methods for making fuels from biomass come in two forms — biological or chemical. The latest approach combines the best of both worlds, and heralds the advent of a second generation of biofuels.

Carbohydrates from biomass will almost certainly provide the source of carbon-based fuels of the future. But the fuel of choice and the method of production are still uncertain^{1,2}. In the absence of an optimal process, there is a vigorous debate over whether the biomass conversion system should be thermochemical (using heat and metal catalysts) or biological (using enzymes and microorganisms). The fuels produced reflect the process involved: ethanol is the product of biological conversion, whereas synthetic diesel (a mixture of saturated hydrocarbons known as alkanes) is that of thermochemical methods. Reporting on page 982 of this issue, Dumesic and colleagues³ blur the lines between these two routes. They apply catalytic techniques to biologically derived sugar molecules to yield a potential fuel called 2,5-dimethylfuran (DMF).

The biggest headache for those developing biofuels is the stark contrast between what we have (biomass rich in carbohydrates) and

what we want (oxygen-deficient fuels). Carbohydrates, such as cellulose and starch, comprise as much as 75% of biomass sources — typically, corn (maize), trees and grasses. These carbohydrates take the form of large polymer chains assembled from thousands of sugar units; each unit contains six carbon atoms and a similar number of oxygen atoms. But optimal fuel molecules for conventional power systems, such as automobile engines, are nothing like plant carbohydrates. They must be small (only 5–15 carbons) and contain little oxygen. The challenge of producing biofuels is finding a way of breaking down long carbohydrate chains to form small, usable molecules, while simultaneously removing the oxygen and minimizing the loss of energy value of the original biomass.

The initial thermochemical approach to this problem was to partially oxidize biomass to produce 'synthesis gas' (a mixture of carbon monoxide and hydrogen), which was then

reacted on metal catalysts to form synthetic diesel (Fig. 1a). In other words, the carbohydrates were broken down to single-carbon atom components and then built up again to larger molecules. But this process generates significant entropy, and about half of the carbohydrate's energy is lost in the process⁴.

The alternative biological approach uses enzymes to break down starch and cellulose to glucose (Fig. 1b). This sugar can then be further processed into fuel molecules using an ever-increasing array of microorganisms and enzymes⁵. The current rapid, worldwide expansion of ethanol-plant construction depends on the hard-working yeast *Saccharomyces cerevisiae*. This organism ferments glucose into two equivalents of ethanol, transferring some of the original carbohydrate oxygens to carbon dioxide. However, the fermentation process is relatively slow (on the order of days), and a better transportation fuel would contain less oxygen, making it more

energy dense and less likely to absorb water.

Dumesic and colleagues' hybrid process³ for converting carbohydrates to DMF provides the benefits of both the older methods while avoiding their pitfalls. It does this by combining new catalytic chemistry with techniques from conventional biological systems (Fig. 1c). The process begins by enzymatically cleaving carbohydrates into fragments, which are rearranged to form the sugar fructose. Rather than fermenting this highly oxygenated sugar, the next step is an acid-catalysed reaction that expels three oxygen atoms, eliminating them as water molecules. This generates an intermediate compound, 5-hydroxymethylfurfural (HMF), which is immediately extracted to prevent further undesired reactions. HMF will probably provide a valuable biologically derived chemical building-block for the future, but its multiple chemical groups and polar nature prevent it from consideration as a fuel⁶. The authors therefore developed a carbon-supported copper–ruthenium catalyst that allowed two more oxygen atoms to be cleaved out of HMF using hydrogen gas, so yielding the desired fuel, DMF.

Compared with existing methods, the benefits of DMF production are clear. Using fructose as the starting material avoids many of the energy-intensive procedures common to thermochemical techniques — for example, the gas-compression step that is necessary to recombine carbon monoxide into synthetic diesel. By replacing biological processes, such as fermentation, with more conventional catalytic methods, the conversion of a sugar to fuel can be hundreds to thousands of times faster than before; this permits the use of much smaller refineries, and could reduce capital investment. The introduction of catalytic techniques could also spawn a second generation of biofuels with improved physical properties. Although it is

too early to determine the fate of DMF as a fuel, its high energy density and hydrophobic nature certainly recommend it as an energy carrier. If it is to be used as a transportation fuel, further examination of its combustion capabilities, as well as its impact on health and the environment, is required.

Combining techniques from diverse areas is a well-known strategy in science and engineering, but Dumesic and colleagues' process³ is a ground-breaking example of interdisciplinary engineering that may well dictate the future of biomass conversion. Their process will no doubt inspire many other combinations of chemical and biological reactions for biofuel production. This work will also focus attention on the potential of other hybrid systems, such as metal catalysts that directly process carbohydrates into fuels^{7,8}, or microorganisms that reassemble synthesis gas to biofuels⁹. Such technologies could completely change current thinking about biofuel production, so that the competition between thermochemical and biological methods ends in a tie. ■

Lanny D. Schmidt and Paul J. Dauenhauer are in the Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Avenue SE, Minneapolis, Minnesota 55455, USA.

e-mail: schmi001@umn.edu

1. Sanderson, K. *Nature* **444**, 673–676 (2006).
2. Ragauskas, A. J. et al. *Science* **311**, 484–489 (2006).
3. Román-Leshkov, Y., Barrett, C. J., Liu, Z. Y. & Dumesic, J. A. *Nature* **447**, 982–985 (2007).
4. Huber, G. W., Iborra, S. & Corma, A. *Chem. Rev.* **106**, 4044–4098 (2006).
5. Stephanopoulos, G. *Science* **315**, 801–804 (2007).
6. Bicker, M., Hirth, J. & Vogel, H. *Green Chem.* **5**, 280–284 (2003).
7. Asadullah, M., Tomishige, K. & Fujimoto, K. *Cat. Comm.* **2**, 63–68 (2001).
8. Salge, J. R., Dreyer, B. J., Dauenhauer, P. J. & Schmidt, L. D. *Science* **314**, 801–804 (2006).
9. Datar, R. P., Shenkman, R. M., Cateni, B. G., Huhnke, R. L. & Lewis, R. S. *Biotechnol. Bioeng.* **86**, 587–594 (2004).

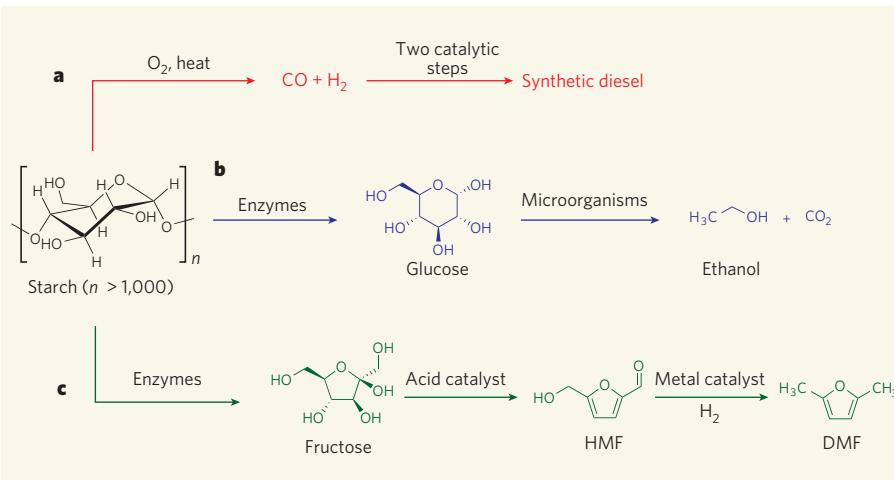


Figure 1 | Conventional and hybrid biofuel production processes. **a**, The conventional thermochemical route to biofuels breaks down starch (or other biomass) into a mixture of carbon monoxide and hydrogen. This mixture is then converted catalytically into synthetic diesel. **b**, Conventional biological routes convert starch to glucose, which is then fermented by microorganisms to produce ethanol. **c**, Dumesic and colleagues³ adopt a hybrid route that enzymatically converts starch into fructose. An acid-catalysed reaction converts the fructose into 5-hydroxymethylfurfural (HMF), which undergoes another catalytic reaction with hydrogen to yield the potential fuel 2,5-dimethylfuran (DMF).



50 YEARS AGO

"Marine creatures" [review of five books] — Far from causing a decline in reading, as was once predicted, it is now becoming evident that television has led to a greatly increased sale of books dealing with topics which have proved popular on the screen. This is perhaps most evident in archaeology, but it is becoming noticeable in other fields too. The growing sport of undersea swimming has reinforced the demand for books about sea life, the publication of which has received a further fillip from the film and television successes of Hans Hass and Jacques Cousteau. We cannot blame the publishers for trying to satisfy this demand, but we can blame them for publishing books seemingly written in haste merely to profit from this fashion.

From *Nature* 22 June 1957.

100 YEARS AGO

"The rainbow" — In "Poems by Two Brothers", written by the Tennysons, and published in 1827, is a poem called "Phrenology." The following lines occur:

"Shall we, with Glasgow's learned Watt, maintain
That you bright bow is not produced
by rain?
Or deem the theory but ill surmised,
And call it light (as Brewster)
polarised?"

Can any of your readers kindly tell me (1) what view was held by James Watt about the rainbow? (2) If Brewster was the first to point out that its light is polarised? Brewster states that he observed the fact in 1812. (3) Having regard to the date (1827), what were the most probable sources of information to which the writer of the poem was indebted? Lord Tennyson kindly informs me that the poem was probably written by Charles Tennyson. Chas. T. Whitmell From *Nature* 20 June 1907.

50 & 100 YEARS AGO

IMMUNOLOGY

Short-term memory

Benjamin N. Gantner and Harinder Singh

Chemical modification of histone proteins can affect the expression of their associated genes. Some immune cells seem to exploit this process to avoid excessive inflammation while fighting invading pathogens.

The innate immune system has several essential roles: it must detect infectious pathogens, initiate antimicrobial mechanisms to remove them and trigger inflammation to activate additional immune responses such as fever. This last function is tricky because too little inflammation will lead to an ineffective response and too much can lead to septic shock and death. So how does the innate immune system prevent excessive responses while repeatedly encountering the same pathogen during an infection?

It has been appreciated for some time that particular innate immune cells — macrophages — can dampen their reactions; however, on page 972 of this issue, Foster *et al.*¹ suggest a more complex change in the sensitivity of these cells to pathogens. They find that macrophages selectively modify the histone proteins that package the genes activated in response to

pathogens, to adapt to repeated exposure.

More than a century ago, Richard Pfeiffer discovered² that components of dead bacteria, which he called endotoxins, could kill test animals; this is now known to be due to an excessive inflammatory response. Endotoxins are highly conserved components of pathogens that are recognized by the innate immune system. Among the most potent endotoxins is lipopolysaccharide (LPS), which is a component of the outer membrane of bacterial cells. It is recognized by Toll-like receptor 4 (TLR4) on the surface of macrophages, where it initiates the molecular signalling pathways that lead to the activation of proinflammatory and antimicrobial genes.

Paul Beeson's seminal work³ in the 1940s uncovered a fascinating twist to the endotoxin response in humans. The typhoid vaccine was used in patients with syphilitic infection of

the nervous system to slow disease progression. Although initial exposure to this vaccine caused fever, repeated daily exposures suppressed the induction of fever and led to a state of tolerance. Similarly, macrophages, which underpin many of the physiological responses to endotoxins, exhibit LPS tolerance on repeated stimulation⁴. Consequently, through limiting the production of proinflammatory molecules, tolerance is thought to provide a mechanism for restraining systemic inflammation and avoiding septic shock⁴.

The induction of a 'tolerized' state in a macrophage on repeated stimulation with LPS involves desensitization of the downstream signalling pathways. In naive macrophages, LPS stimulation activates signalling pathways mediated by the MAPK and IKK enzymes. These, in turn, trigger gene transcription factors AP-1 and NF-κB respectively to activate the expression of both antimicrobial and proinflammatory genes⁴. On restimulation with LPS, the signalling response of tolerized macrophages is attenuated. Several proteins, including IRAK-M, SOCS-1 and SHIP, which are components of negative-feedback loops induced by these signalling pathways, constrain LPS signalling during the initial stimulation and play a role in establishing tolerance⁵. So, based on this conceptual framework, repeated LPS sensing by macrophages should impair the induction of both proinflammatory and antimicrobial genes, thereby compromising the innate immune response to various pathogens. This is not the case.

Foster and colleagues¹ hypothesized that macrophages may have gene-specific control mechanisms to circumvent the problem posed by generalized desensitization to reiterative LPS signalling. Accordingly, genes encoding proinflammatory mediators should be transiently inactivated in tolerized macrophages, whereas those encoding antimicrobial proteins should remain inducible on repeated LPS stimulation.

They performed genome-wide expression analysis of naive and tolerized macrophages, and, indeed, identified two classes of genes — tolerizable and non-tolerizable. On restimulation with LPS, tolerizable genes, largely proinflammatory genes such as *IL-1* and *IL-6*, were not induced. Non-tolerizable genes were expressed to the same, or even greater, level compared with the initial stimulation; this class largely contained antimicrobial genes.

The authors also show that, on initial LPS stimulation, the histones associated with both tolerizable and non-tolerizable genes transiently undergo chemical modifications that promote the expression of their associated genes. As expected, RNA polymerase II — the enzyme that transcribes genes into messenger RNA — is also initially recruited to both classes of gene. However, on restimulation, tolerizable genes selectively failed to undergo activating histone modifications or to recruit RNA polymerase II (Fig. 1). This refractory state

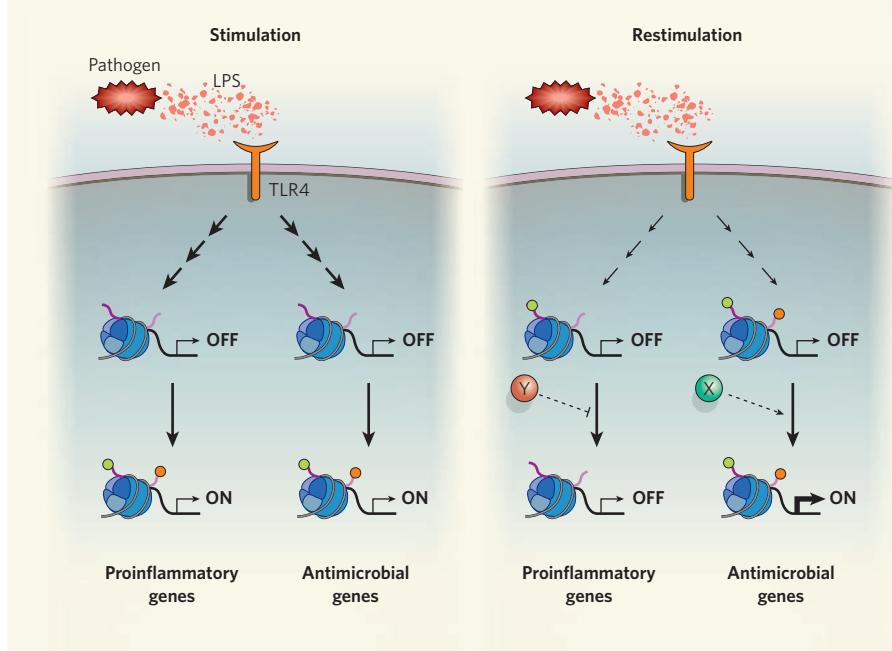


Figure 1 | Chromatin modifications and the response of macrophages to pathogens. In response to bacterial lipopolysaccharide (LPS), a TLR4-mediated signalling cascade is activated in macrophages, which induces the expression of both antimicrobial and proinflammatory genes. However, on restimulation with LPS — which is accompanied by desensitization of the downstream signalling pathways — antimicrobial genes are selectively activated. Foster *et al.*¹ show that this is in part mediated by chromatin modifications (green and orange balls), which selectively mark the antimicrobial genes, among others. Consequently, on restimulation — and with the help of an unknown activator (X) — the expression of antimicrobial genes is increased to the same or higher levels. By contrast, certain other genes, including proinflammatory genes, lose their histone marks after initial stimulation, preventing their expression on restimulation; this is also aided by an unknown repressor (Y). This process prevents excessive inflammation, which can be deadly.

was reversed by inhibitors of the enzymes that remove activating histone modifications, indicating a role for these enzymes in turning off tolerizable genes.

It remains possible that these inhibitors have additional effects unrelated to histone modification. Therefore more definitive evidence will be required, such as the identification of the molecules required for gene 'silencing'. Moreover, as loss of negative regulators of TLR4 signalling can also result in reactivation of proinflammatory genes⁵, it is likely that macrophages acquire tolerance by self-reinforcing mechanisms of attenuated TLR4 signalling, as well as repressive histone modifications.

Foster *et al.*¹ found that, in contrast to tolerizable genes, the chromatin of non-tolerizable genes maintained some activating histone modifications after the initial LPS stimulation. This is likely to prime the non-tolerizable genes for even more robust activation on subsequent LPS stimulations. This possibility is supported by the more rapid and vigorous induction of non-tolerized genes on LPS restimulation, despite the attenuated TLR4-mediated signalling.

The establishment of primed and silenced chromatin states of non-tolerizable and tolerizable genes respectively seems to depend on *de novo* synthesis of one or more factors as a consequence of the initial LPS signal (X and Y; Fig. 1). The nature and mechanisms of action of these essential secondary regulators are unknown. The β-arrestin proteins are potential candidate regulators of non-tolerizable gene expression; they not only attenuate TLR-induced signalling but also bind to transcriptional complexes and promote activating histone modifications⁶. This raises the exciting possibility that, in macrophages, β-arrestins could both restrain TLR signalling and maintain the robust expression of the non-tolerizable antimicrobial genes.

Another mechanism for sustaining and augmenting the expression of non-tolerizable genes on repeated LPS stimulation is suggested by a study in yeast⁷. The sugar galactose induces the expression of yeast *GAL* genes, which encode enzymes required for its usage. An initial exposure to galactose primes these genes for stronger activation on a secondary exposure — a phenomenon termed transcriptional memory. This is essentially dependent on the replacement of the canonical histone protein, H2A, which is associated with the *GAL* gene promoters, with a variant histone called H2A.Z. It will be interesting to determine whether, in LPS-stimulated macrophages, the priming of non-tolerized genes also involves histone replacement.

The discovery of Toll-like receptors greatly enhanced our understanding of immunology by establishing that the innate immune system also uses recognition systems with specificity for molecules displayed by pathogens — similar to the antigen receptors of T and B cells. Stimulation of antigen receptors activates

lymphocyte differentiation programmes that permanently modify the adaptive immune response by generating memory B and T cells. The findings of Foster *et al.*¹ indicate that, after their receptor-mediated triggering, cells of the innate immune system are also capable of undergoing a differentiation programme with features of memory, albeit one that is short term. It is perhaps not surprising that many of the hallmarks of the adaptive immune system are manifest in the evolutionarily ancient innate immune system in a more primitive form. ■

Benjamin N. Gantner and Harinder Singh are in the Department of Molecular Genetics and Cell Biology, Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637, USA. e-mail: hsingh@uchicago.edu

1. Foster, S. L., Hargreaves, D. C. & Medzhitov, R. *Nature* **447**, 972–978 (2007).
2. Beutler, B. & Rietschel, E. T. *Nature Rev. Immunol.* **3**, 169–176 (2003).
3. Van Epps, H. L. *J. Exp. Med.* **203**, 1137 (2006).
4. Dobrovolskaia, M. A. & Vogel, S. N. *Microbes Infect.* **4**, 903–914 (2002).
5. Beutler, B. *Immunity* **21**, 134–135 (2004).
6. Ma, L. & Pei, G. *J. Cell Sci.* **120**, 213–218 (2007).
7. Brickner, D. G. *et al.* *PLoS Biol.* **5**, e81 (2007).

MATERIALS SCIENCE

Reflections on ionic liquids

Robin D. Rogers

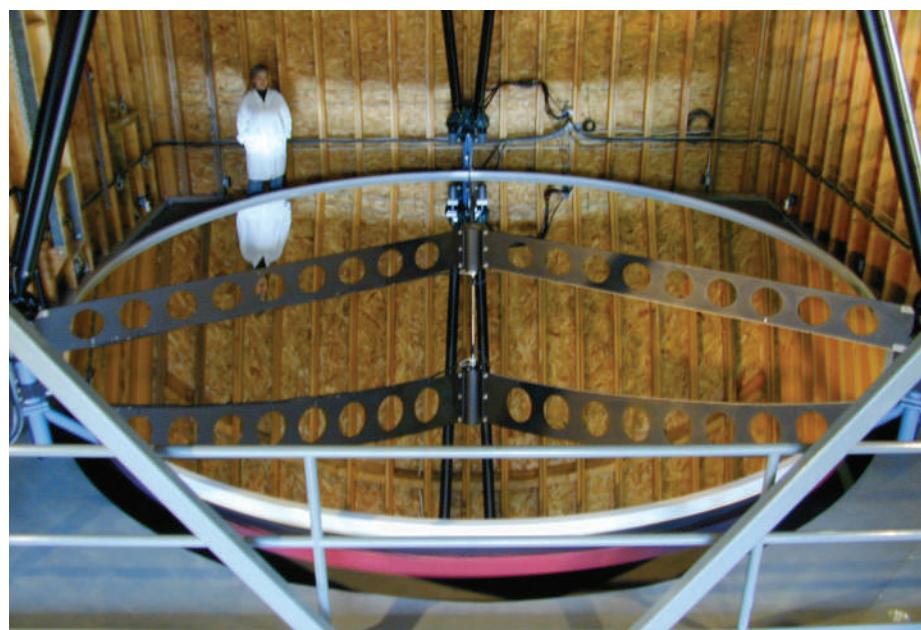
Ionic liquids are generally regarded as solvents, but these modular, tunable compounds have far greater technological potential. With a coat of silver, they become ideal materials for the liquid mirror of a space telescope.

Ionic liquids seem to defy common sense. Most ionic compounds are crystalline solids with high melting points, but these fascinating salts melt at temperatures below 100 °C; indeed, many are liquids at room temperature. Their melt forms are composed of discrete cations and anions¹ that can be individually customized, allowing the synthesis of a wide range of liquid materials with tunable physical, chemical and biological properties.

There are thought to be about a million possible pure ionic liquids, and 10¹⁸ ternary liquid mixtures, so anyone designing the

perfect liquid material for a given application has a lot of room for manoeuvre. A particularly striking example is described by Borra *et al.*² on page 979 of this issue. They have coated an ionic liquid with colloidal silver particles, yielding a material that could be used as a liquid mirror in a telescope.

Ionic liquids are not new³, but they have recently received intense worldwide scrutiny as possible environmentally friendly solvents⁴ because many are non-volatile. This has fuelled a technological revolution, powered by the sheer number of unstudied liquids that might



P. HICKSON, UNIV. COLUMBIA

Figure 1 | Liquid-mirror telescopes. The Large Zenith Telescope is sited in the Malcolm Knapp Research Forest, 70 kilometres east of Vancouver, Canada. It uses the world's largest liquid mirror — a smoothly spinning pan of mercury, 6 metres in diameter. Borra *et al.*² describe the preparation of a reflective ionic liquid that could also be used as a liquid mirror.

be fine-tuned for specific purposes (although the 'green' credentials of ionic liquids have been questioned by reports that some of these compounds are toxic⁵). It was, therefore, inevitable that new applications would emerge from the growing number of scientific and technological disciplines studying these liquids.

Ionic liquids are known for their distinct physical properties (such as low or non-volatility, thermal stability and large ranges of temperatures over which they are liquids⁶), chemical properties (such as resistance to degradation, antistatic behaviour, chirality and high energy density) and biological activities (such as antimicrobial and analgesic properties⁷). But what is less appreciated is that these properties in individual ionic liquids can be combined in composite materials to afford multifunctional designer liquids. It is therefore refreshing to see a study² that focuses on the unique attributes and uses of ionic liquids, rather than on whether they are green or toxic.

Borra *et al.*² use an ionic liquid to solve a problem in making liquid mirrors for telescopes (Fig. 1). Liquid mirrors have several advantages over traditional mirrors for such applications — for example, they have excellent optical properties and their surfaces form perfectly smooth parabolas. It has been proposed that a telescope on the Moon with a large liquid mirror (20–100 metres across) could provide unprecedented views of deep optical fields, so advancing our knowledge of the early Universe.

A major roadblock to the implementation of a liquid-mirror telescope is finding a stable liquid support for the reflective coating that can resist the extreme environment of space. Thus, the support must have high viscosity, a very low melting point or glass transition temperature, and no vapour pressure. Borra *et al.*² used vacuum vaporization to coat silver onto several liquids, including silicone oil, a block copolymer and an ionic liquid. Of the liquids tested, the ionic liquid came closest to having the desired physical properties, and also yielded the most reflective material with a stable coating of silver. Furthermore, the coating process could be improved by depositing chromium on the ionic liquid before the silver, and provided a surface with even better optical quality than silver alone. Further improvements to the ionic liquid will be necessary before it can be used in a space telescope. Nevertheless, this report² surpasses most descriptions of these liquids because the application depends completely on the physical and chemical properties of the ionic liquid — in fact, it seems that only an ionic liquid will do.

The approach taken by Borra *et al.*² was first to define the properties needed for an ideal liquid-mirror support, and then to identify an ionic liquid as being suited for that purpose. They focused mainly on the physical properties of the liquid, but its chemical properties should also be carefully considered — for example, the solubility and reactivity of the reflecting metal

(or metallic colloid) with the liquid. Such considerations may lead to improved methods of metal deposition, or to new forms of liquid mirrors.

One problem for the future is finding exactly the right ionic liquid for the job, even though the properties required for a liquid-mirror material are known. Given the vast number of possible ionic liquids to choose from, and the fact that few rules exist for customizing them (other than rules of thumb), the selection of an appropriate ionic liquid is arduous and often hit-and-miss. Anyone developing ionic liquids for technological applications faces this challenge, and there is always the danger that a competitor will chance upon a better choice. Hope lies in the major efforts now being made to model and predict the properties of ionic liquids, although such predictive methods will take time to develop.

In the meantime, a knowledge base of interdisciplinary data is rapidly being generated

for ionic liquids. This should fuel innovative ideas and applications that will take these liquids far beyond the realm of mere solvents. The idea that ionic liquids could pave the way for exciting fundamental science has yet to be recognized. Nonetheless, the potential power of these materials is clear: one need only look in the mirror. ■

Robin D. Rogers is in the Department of Chemistry and Center for Green Manufacturing, The University of Alabama, Tuscaloosa, Alabama 35487, USA.

e-mail: rdrogers@bama.ua.edu

- Wasserscheid, P. & Welton, T. (eds) *Ionic Liquids in Synthesis* (Wiley-VCH, Weinheim, 2003).
- Borra, E. F. *et al.* *Nature* **447**, 979–981 (2007).
- Walden, P. *Bull. Acad. Sci. St Petersburg* 405–422 (1914).
- Fremantle, M. *Chem. Eng. News* **76** (30 March), 32–37 (1998).
- Nature* 10.1038/news051031-8 (2005).
- Deetlefs, M., Seddon, K. R. & Shara, M. *Phys. Chem. Chem. Phys.* **8**, 642–649 (2006).
- Pernak, J., Sobaszkiewicz, K. & Mirska, I. *Green Chem.* **5**, 52–56 (2003).

EVOLUTIONARY BIOLOGY

Re-crowning mammals

Richard L. Cifelli and Cynthia L. Gordon

The evolutionary history of mammals is being tackled both through molecular analyses and through morphological studies of fossils. The 'molecules versus morphology' debate remains both vexing and vibrant.

On page 1003 of this issue, Wible and co-authors¹ announce the discovery of a well-preserved mammal from Mongolia dated at between 71 million and 75 million years old. The fossil, dubbed *Maelestes gobiensis*, is noteworthy in its own right: finds of this sort are exceptional in view of the generally poor record of early mammals.

More interesting, though, is what this fossil and others from the latter part of the age of dinosaurs (the Cretaceous period, about 145 million to 65 million years ago) have to say about the rise of mammalian varieties that populate Earth today. The authors have gone much further than describing an ancient fossil specimen, and present a genealogical tree depicting relationships among the main groups of living and extinct mammals. Here, all Cretaceous fossil mammals are placed near the base of the tree, as dead 'side branches', well below the major tree 'limbs' leading to living mammals. These results differ strikingly from those of other recent palaeontological studies^{2,3}.

Chronologically speaking, this new analysis¹ is eye-popping because it places direct ancestry of today's mammals near the Cretaceous-Tertiary (K/T) boundary about 65 million years ago. This is much younger than dates based on molecular biology — for example, a recent and comprehensive analysis by Bininda-Emonds *et al.*⁴ pushed that ancestry back more than

twice as far into the geological past, to some 148 million years ago. The conflicting results of these palaeontological¹ and molecular⁴ studies have profound implications for understanding the evolutionary history of mammals, and for understanding the pace and nature of evolution generally.

Three main groups of living mammal are recognized: the egg-laying monotremes such as the platypus; marsupials (kangaroos, koalas, opossums and so on); and placentals, which constitute the most varied and diverse group, including everything from bats to whales and accounting for more than 5,000 of the 5,400 or so living mammals. Fossils can be placed within one of these three 'crown' groups only if anatomical features show them to be nested among living species⁵.

The placental crown group, which is of primary interest here, represents the living members of a more encompassing group, Eutheria, which includes extinct allied species, the oldest of which dates to about 125 million years ago⁶. Herein lies a central problem: because of inadequate preservation and/or non-comparability with living species, the affinities of many early mammals have been contentious. Certain Cretaceous fossils have been previously recognized as members of the placental crown group; some analyses suggest the presence of placental superorders in the Cretaceous^{2,3}, but referral of

such ancient fossils to living orders is dubious⁵. For context, placentals encompass four major divisions, or superorders, each containing one to six orders, such as Cetacea (whales), Primates and Rodentia.

The study by Wible *et al.*¹ is ground-breaking because it brings a wealth of new data into play: it includes every informative Cretaceous fossil and is based on comparison of more than 400 anatomical features. Palaeontologically, the authors' evolutionary tree is iconoclastic in demoting many previously recognized members of the placental crown group to the status of 'stem' species, or generalized eutherians. In this scheme, the oldest-known placental is a rabbit-like mammal from Asia, dated to about 63 million years ago.

Of more general interest are the implications of this tree for dating mammalian evolutionary radiations, and the factors that may have affected them. Following extinction of non-avian dinosaurs at the K/T boundary, the fossil evidence shows that eutherians underwent significant radiations in the Palaeocene (between 65 million and 55 million years ago), and that most of the modern groups appeared and flourished later. One cannot help but notice an analogy between this 'bushy' radiation and the initial explosion of complex life-forms some 500 million years ago. In both cases, the explosion is followed by the extinction of lineages

that presumably represent failed evolutionary experiments, with the concomitant emergence and radiation of modern types⁷.

By coincidence, the appearance of Wible and colleagues' paper¹ comes hard on the heels of that by Bininda-Emonds *et al.*⁴, which was published in March. The two studies — one based on anatomy (emphasizing fossils) and the other on molecular biology (living species only) — come to very different conclusions about the timing of mammalian evolution. As such, they represent the latest volleys in the 'molecules versus morphology' debate⁵.

Previous studies have identified three models for the origin and diversification of placental mammals⁸: 'explosive', in which divergence of most superorders and orders occurred near and following the K/T boundary; 'long fuse', differing in the significantly earlier diversification of superorders; and 'short fuse', which calls for diversification of both superorders and orders well back in the Cretaceous. The study by Bininda-Emonds *et al.*⁴, which integrates results of about 2,500 subtrees that collectively include 99% of living mammal species, is the most comprehensive of its kind to date⁹. It yields support for both the short-fuse (groups including at least 29 living species) and the long-fuse (less diverse groups) models, with a lull in diversification of placentals *per se* following the K/T boundary (Fig. 1a).

By contrast, Wible and colleagues' morphological work¹ strongly supports the explosive model (Fig. 1b).

These results^{1,4} show a widening rather than a narrowing of the gap between the conclusions drawn from morphological and molecular studies. Why the difference? The two studies are based on independent lines of evidence, each with its own shortcomings. The fossil record is notorious for its incompleteness, thereby leaving open the possibility of new discoveries that radically alter the picture. Some studies suggest, however, that the existing fossil record is complete enough to be taken at face value^{8,10}. The principal issue with molecular studies has to do with assumptions about the 'molecular clock' and variations in the rates of gene substitution on which such research is based. Yet there are also important points of congruence among the results, notably in the geometry of the evolutionary trees, suggesting that neither type of data has an exclusive claim to validity.

Where do we go from here? For palaeontologists, the answer lies in filling the gaps in the fossil record. One new fossil, such as a Cretaceous giraffe, could send Wible and co-authors scrambling back to the drawing-board. And those involved in molecular studies must continue to develop more sophisticated methods to account for gene-substitution rates that vary according to lineage, geological time interval, body size and other factors¹¹.

For the onlooker, however, the big question is whether the floodgates of mammalian evolution were ecologically opened by dinosaur extinctions at the K/T boundary. The answer seems to be 'yes', at least in part¹². Evolutionary trees are essential, but further levels of analysis are needed to interpret changes in terrestrial ecosystems and the assemblages of mammals that inhabited them. In this context, perhaps attention has been too narrowly focused on crown placentals^{4,9}: other eutherians, marsupials and mammalian varieties were also present during this exciting time. Ultimately, interpreting the dynamics of mammalian evolution will depend on integrating genealogical investigations — both palaeontological and molecular — with complementary studies of palaeoecology and of the role that each species played in its respective community.

Richard L. Cifelli is at the Sam Noble Oklahoma Museum of Natural History, 2401 Chautauqua Avenue, Norman, Oklahoma 73072, USA.

Cynthia L. Gordon is in the Department of Zoology, University of Oklahoma, Norman, Oklahoma 73019, USA.

e-mails: rlc@ou.edu; cindyg@ou.edu

- Wible, J. R., Rougier, G. W., Novacek, M. J. & Asher, R. J. *Nature* **447**, 1003–1006 (2007).
- Archibald, J. D., Averianov, A. O. & Ekdale, E. G. *Nature* **414**, 62–65 (2001).
- Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs: Structure, Relationships, and Paleobiology* (Columbia Univ. Press, New York, 2004).
- Bininda-Emonds, O. R. *et al.* *Nature* **446**, 507–512 (2007).

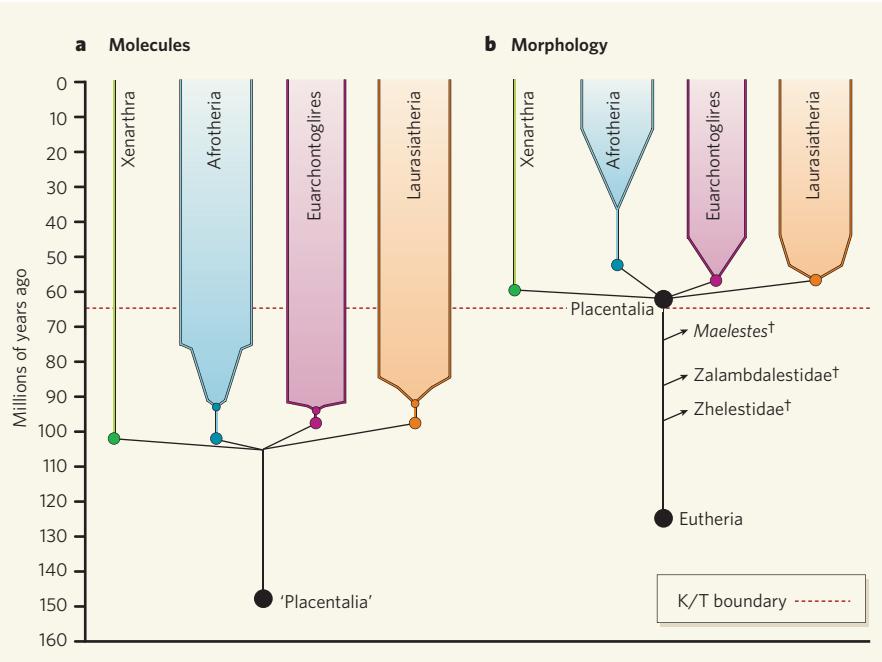


Figure 1 | Two views (simplified) of the diversification of the major orders of modern placental mammals. **a**, The picture provided by the molecular analyses of Bininda-Emonds *et al.*⁴. In this, inter-ordinal diversification of the four main placental superorders occurred in the mid-Cretaceous, with intra-ordinal diversification happening soon thereafter (although this is not the case for all lineages). 'Placentalia' is equivalent to Eutheria, as used elsewhere^{1,8}. **b**, The picture arising from the morphological (fossil) studies of Wible *et al.*¹. Here, the modern orders of placentals did not appear and diversify until after the K/T boundary, with many Cretaceous mammals (such as *Maelestes*[†]) being relegated to evolutionary dead-ends. These fossils near the base of the tree are included in the broader group Eutheria, whose living representatives are the placentals. The placental superorders are the Xenarthra (sloths and armadillos, for example), Afrotheria (elephants, sea cows), Euarchontoglires (primates, bats, rodents) and Laurasiatheria (whales, carnivores, shrews). No genealogical relationships are implied in either tree. †, extinct group.

5. Benton, M. J. *BioEssays* **21**, 1043–1051 (1999).
6. Ji, Q. *et al.* *Nature* **416**, 816–822 (2002).
7. Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History* (Norton, New York, 1989).
8. Archibald, J. D. & Deutscher, D. H. *J. Mammal. Evol.* **8**, 107–124 (2001).
9. Penny, D. & Phillips, M. J. *Nature* **446**, 501–502 (2007).
10. Foote, M., Hunter, J. P., Janis, C. M. & Sepkoski, J. J. *Science* **283**, 1310–1314 (1999).
11. Springer, M. S., Murphy, W. J., Eizirik, E. & O'Brien, S. J. *Proc. Natl Acad. Sci. USA* **100**, 1056–1061 (2003).
12. Wilson, G. P. Thesis, Univ. California (2004).

BIOPHYSICS

Proteins hunt and gather

David Eliezer and Arthur G. Palmer III

Some proteins do not fold fully until they meet their functional partners. Folding in concert with binding allows an efficient stepwise search for the proper structure within the final complex.

In higher organisms, many proteins, including some involved in critical aspects of biological regulation and signal transduction, are stably folded only in complex with their specific molecular targets. On page 1021 of this issue, Sugase *et al.*¹ elucidate a three-step mechanism by which one such ‘intrinsically disordered’ protein binds to its cognate folded protein target. This mechanism indicates a bipartite strategy for this class of protein in optimizing the search for partner molecules. An initial encounter complex, formed through weak, nonspecific interactions, facilitates the formation of a partially structured state, which makes a subset of the final contacts with the target. This intermediate conformation allows an efficient search for the final structure adopted by the high-affinity complex.

Previous work by these authors^{2,3} described the conformational preferences of an intrinsically disordered polypeptide that constitutes part of the gene transcription factor, CREB; this polypeptide is known as the phosphorylated kinase inducible activation domain (pKID). When found in a high-affinity complex with the KIX domain of the CREB-binding protein, pKID forms two α -helices (A and B) in its amino- and carboxy-terminal regions, respectively. Helix B makes intimate contacts with a hydrophobic groove on the KIX surface, whereas helix A forms a less extensive interface with KIX (ref. 2). In the absence of KIX, pKID is largely, but not completely, disordered. Its amino-terminal region intermittently forms helix A, but its carboxy-terminal region is more unstructured^{3,4}.

The different molecular species formed during pKID binding to KIX interconvert kinetically; hence, neither the encounter complex nor the intermediate complex can be isolated and studied directly. To characterize these species, Sugase and colleagues used techniques that rely on the exquisite sensitivity of the resonance frequencies observed in nuclear magnetic resonance (NMR) spectroscopy. In particular, time-dependent changes in local chemical environments and molecular

structures modify the resonance frequencies by altering the magnetic fields experienced by individual atomic nuclei.

The specific effects of environmental and structural changes on NMR spectra depend on whether the kinetic transition rate constants linking different molecular states are larger than, comparable to or smaller than the differences in the resonance frequencies of these states. These three regimes are termed fast, intermediate and slow exchange, respectively. For example, in the fast-exchange limit, the observed frequency of a resonance signal (which in NMR spectroscopy is called the chemical shift) is the population-weighted average of individual resonance frequencies for different states. The width of the resonance signal (which is proportional to the transverse relaxation rate constant for the nuclear magnetization) depends on the variation in individual resonance frequencies and on the transition rates.

The approach developed by Sugase *et al.* will probably be widely applicable to the study of other protein–protein binding reactions. Using established techniques, known as ^1H – ^{15}N single-quantum correlation (HSQC) and ^{15}N transverse relaxation dispersion, the authors monitored changes in chemical shifts and relaxation rate constants as a function of the concentration ratio of the two interacting proteins.

The HSQC technique yields highly sensitive and well-resolved NMR spectra that allow detailed monitoring of the chemical shifts for the ^1H and ^{15}N nuclei of amide groups in proteins. The relaxation dispersion experiment measures the transverse relaxation rate constants for the amide ^{15}N nuclei in the presence of applied radiofrequency fields, as strong effective fields partially suppress the relaxation caused by transitions between molecular states with different resonance frequencies. These two techniques allow the identification and structural characterization of weakly populated, or rare, conformational states that arise during coupled binding and folding processes. They also allow quantification of

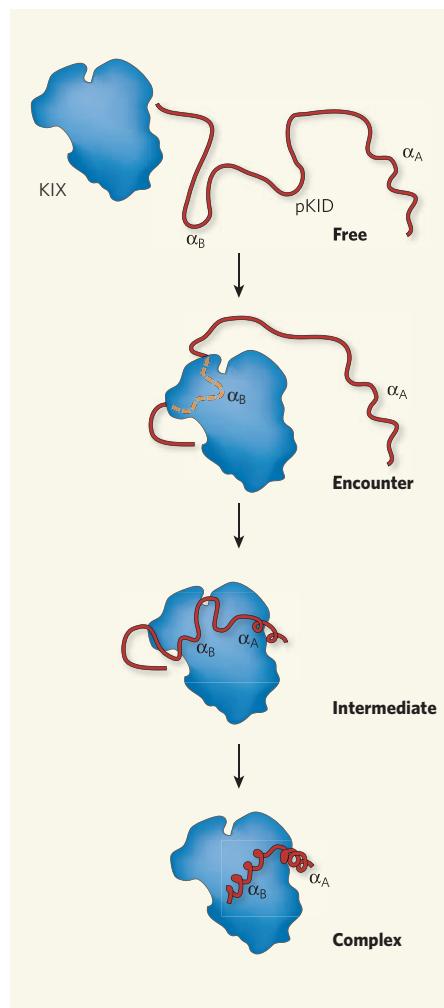


Figure 1 | A complex encounter between disorder and order. Interaction between the pKID domain of the gene transcription factor CREB and the KIX domain of the CREB-binding protein occurs in the cell nucleus to regulate gene expression. By elucidating the three-step binding reaction between pKID and KIX using NMR spectroscopy, Sugase *et al.*¹ identified four states along the reaction pathway. Initially, the highly disordered, free state of pKID partially populates helix A (α_A). In the encounter complex with KIX, pKID is tethered by nonspecific hydrophobic contacts in its helix B region (α_B). The intermediate state is characterized by a specifically bound and largely configured helix A. Finally, in the high-affinity, bound conformation, both helices are fully structured.

the kinetic rate constants linking the different steps along the reaction pathway.

The HSQC spectra of ^{15}N -labelled pKID revealed continuous changes in ^1H and ^{15}N chemical shifts during titration with sub-equivalent quantities of KIX (1:0 to 1:0.5 pKID:KIX concentration ratios). This observation indicates a fast-exchange, reversible interaction between the two proteins, which was confirmed by competition with another peptide that binds to KIX and by mutation of a key amino-acid residue in KIX. The NMR spectrum that is predicted by extrapolating the chemical-shift changes to a 1:1 ratio of these

proteins does not reproduce the spectrum observed for the slow-exchange, high-affinity complex. This result implies the existence of an encounter complex, in which pKID remains largely disordered while residues in its carboxy terminus make initial contacts with KIX (Fig. 1).

Relaxation dispersion data for ¹⁵N-labelled pKID recorded for pKID:KIX ratios ranging from 1:0.95 to 1:1.10 yielded similar kinetic on-rates, or binding rates, for all residues, but different off-rates, or release rates, for clusters of residues in helices A and B. Complexed and free pKID are in slow exchange; therefore, these results can be explained only by the presence of an intermediate state — one that is different from either the encounter or the fully bound complex.

Chemical shifts for the intermediate state extracted from relaxation dispersion showed that the amino-terminal region of pKID almost fully populates the structured helix A conformation, whereas the carboxy-terminal region of the molecule is still searching for its optimal fit with the KIX domain. Full

stabilization of helix B seems to be the ultimate step leading to the final complex.

Interactions between pKID and KIX in the encounter complex are largely nonspecific. They involve the most hydrophobic groups in the highly disordered carboxy-terminal region of pKID, rather than the transiently structured helix A, which is required for high-affinity binding⁵, but apparently only in subsequent steps. In this regard, coupled folding and binding of pKID to KIX resembles the unimolecular folding of globular proteins, which is driven by an interplay between nonspecific hydrophobic interactions, hydrogen-bond formation and association of secondary-structure elements.

Initial, weak, nonspecific interactions, either short or long range, are believed to enhance binding kinetics between well-folded proteins by constraining the diffusional search for a specific binding site. The results of Sugase *et al.*¹ indicate that this mode of action extends to binding events involving intrinsically disordered proteins. Moreover, coupled folding and binding may further restrict diffusional

search within partially structured, tethered intermediate states. Notably, this mechanism implies a stepwise reduction in configurational entropy as energetically favourable interactions are formed. Future applications of this experimental strategy should reveal whether this is a general mechanism in interactions between disordered proteins and their more structured partners. ■

David Eliezer is in the Department of Biochemistry and the Program in Structural Biology, Weill Cornell Medical College, New York, New York 10021, USA. Arthur G. Palmer III is in the Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA.

e-mails: dae2005@med.cornell.edu; agp6@columbia.edu

1. Sugase, K., Dyson, H. J. & Wright, P. E. *Nature* **447**, 1021–1025 (2007).
2. Radhakrishnan, I. *et al.* *Cell* **91**, 741–752 (1997).
3. Radhakrishnan, I., Perez-Alvarado, G. C., Dyson, H. J. & Wright, P. E. *FEBS Lett.* **430**, 317–322 (1998).
4. Hua, Q. X., Jia, W. H., Bullock, B. P., Habener, J. F. & Weiss, M. A. *Biochemistry* **37**, 5858–5866 (1998).
5. Parker, D. *et al.* *Mol. Cell* **2**, 353–359 (1998).

STRUCTURAL BIOLOGY

ESCRT service

Steven L. Alam & Wesley I. Sundquist

The sorting and degradation of cell-surface proteins are essential for cellular homeostasis. The ESCRT-I complex is known to be involved in these events, and new structural findings elucidate its core architecture.

The compartmentalization of cells in higher organisms has brought about the need for reliable trafficking systems to transport cargoes in and out of different membrane-enclosed organelles. The endosomal sorting complex required for transport-I (ESCRT-I), which is evolutionarily conserved from yeast to humans, is one component of the intracellular trafficking machinery. This cytoplasmic complex is required for both the formation of specialized membrane-vesicle-containing compartments known as multivesicular bodies (MVBs) and the budding of enveloped RNA viruses such as HIV. Reporting in *Cell*, Kostelansky *et al.*¹ present the crystal structure of the yeast ESCRT-I core, thereby revealing the structural organization of this important membrane-protein-trafficking complex.

Within a cell, membrane components are frequently shuttled between organelles inside lipid vesicles that bud out of one organelle into the cytoplasm, shed their protein coat — for example, a clathrin coat — and then fuse with another organelle. The mechanisms underlying vesicle formation in these classical trafficking systems are clear². By contrast, the mechanisms of cargo delivery into an organelle's interior or out of a cell, which require the formation

of transport vesicles that bud in the opposite direction — that is, away from the cytoplasm — are much less well understood. The biogenesis of such ‘inward-budding’ vesicles must involve fundamentally different mechanisms, because the machinery that mediates their formation needs to bend the membranes away from the cytoplasm and help pinch off the vesicles from inside the bud neck.

The ESCRT pathway, which carries transmembrane proteins from the outer membrane into the interior of MVBs (Fig. 1, overleaf), is the best-characterized inward-budding system. It helps to regulate the levels of membrane proteins — such as growth-factor receptors — that get tagged for degradation with a small protein known as ubiquitin. Ubiquitinated proteins are sorted into vesicles within MVBs, and are then degraded by the fusion of MVBs with lysosomes. Enveloped RNA viruses also usurp the ESCRT pathway to bud from the cell surface, apparently because the formation of vesicles within MVBs and the budding of these viruses are topologically equivalent³. ESCRT-I is the first dedicated complex in the ESCRT pathway, and therefore the findings of Kostelansky *et al.*¹ are a big step forward in our understanding of both MVB

biogenesis and enveloped-virus budding.

In yeast, the ESCRT-I complex is a heterotetramer consisting of four protein subunits (Vps23, Vps28, Vps37 and Mvb12)^{1,3–5}. There are numerous human variants, each of which has four equivalent subunits⁶. As a result, the core structure of yeast ESCRT-I is a good model for the basic architecture of this complex in higher organisms.

The ESCRT-I core is highly elongated, with a globular ‘headpiece’ sitting at one end of a long thin ‘stalk’ (Fig. 2, overleaf). The headpiece is organized into three almost identical, anti-parallel, helical hairpin structures contributed by conserved segments of Vps23, Vps28 and Vps37 (refs 1,7,8). The Mvb12 subunit contributes a short amino-terminal helix to the headpiece, and interacts with Vps23 to form a short β-sheet that connects the headpiece and the stalk.

The stalk is composed of segments of Vps23, Vps37 and Mvb12 (but not Vps28), and is about 130 Å long, but only about 20 Å wide. The Vps23 and Vps37 subunits form extended helices that pack together with Mvb12 to form a three-helix bundle over one segment of the stalk and a ‘hybrid’ structure over the remaining segment (Fig. 2). The authors found that a basic, amino-terminal, helical extension on Vps37 binds weakly to lipids, and so may help anchor ESCRT-I to membranes.

The results of previous studies^{3–5} had shown that two additional domains not present in the current structure extend from the core, connecting ESCRT-I to ubiquitinated cargoes and to two other ESCRT complexes — the upstream ESCRT-0 and the downstream ESCRT-II. The carboxyl terminus of Vps28 interacts with the Vps36 subunit of ESCRT-II, and the amino

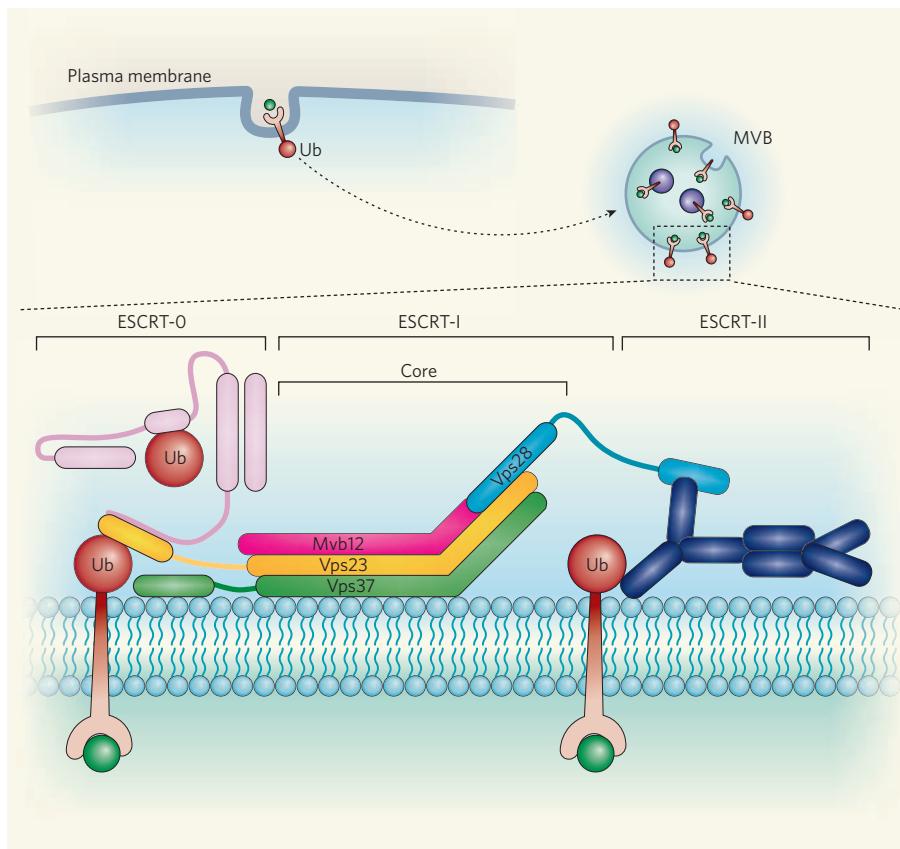


Figure 1 | ESCRT-I and membrane traffic. After internalization, transmembrane proteins tagged with ubiquitin (Ub) enter specialized vesicles called multivesicular bodies (MVBs). The sorting of these proteins to vesicles in MVBs — and their subsequent degradation in lysosomes — is mediated by ESCRT complexes. Kostelansky *et al.*¹ report the core structure of one of these ESCRT complexes, ESCRT-I, which consists of four subunits: Vps23, Vps28, Vps37 and Mvb12. The structures of the two protein-binding domains of this complex have been reported previously^{3–5}, and indicate that ESCRT-I interacts with ESCRT-0 and ESCRT-II. From its core structure, ESCRT-I seems also to interact with lipids on the membrane of MVBs through its Vps37 subunit.

terminus of Vps23 forms complexes with ubiquitin and with proline-rich elements found in the Vps27 subunit of ESCRT-0 and p6^{Gag} proteins of HIV, among others. Thus, we now have high-resolution models for almost the entire ESCRT-I complex and three of its essential interactions.

Although the biological rationale for the highly elongated ESCRT-I conformation is not fully understood, this unusual structure presumably reflects strong functional constraints, particularly as the stalk and the headpiece are both essential for efficient cargo sorting within MVBs¹. Kostelansky *et al.* argue

that the long axis of ESCRT-I is oriented parallel to the membrane plane, implying that the ubiquitin-binding domains of ESCRT-I and ESCRT-II are probably separated by more than 100 Å (Fig. 1).

The ubiquitin-binding domains of ESCRT-I interact with ubiquitin modifications on the cytoplasmic tails of transmembrane cargo proteins, and are separated from similar ubiquitin-binding domains on the ESCRT-II complex by the intervening rod-like ESCRT-I core. This separation probably prevents the sequential passage of ubiquitinated cargoes from one ESCRT complex to the next, assuming that no major conformational changes occur. Instead, the different ubiquitin-binding domains probably function as independent ‘nodes’ within an ESCRT net, concentrating ubiquitinated cargoes into specific regions of the membrane, much as adaptor proteins concentrate cargoes for incorporation into clathrin-coated vesicles².

The 1:1:1:1 ratio of the four subunits in the ESCRT-I heterotetramer — seen both in solution^{1,9–11} and in the new crystal structure¹ — does not match a proposal¹² that the Mvb12 subunits promote ESCRT-I trimerization, and negatively regulate ESCRT-II recruitment. That said, regulated recruitment and co-assembly of ESCRT complexes on cellular membranes must reflect some combination of changes in lipid binding, protein–protein interactions, conformations, oligomeric states and modifications following protein translation^{3–5}. It is therefore important to determine how endogenous ESCRT complexes assemble on native membranes.

The availability of the new ESCRT-I core structure means that structural information is now available for all of the main known components of the ESCRT pathway^{1,3–5}. The challenge now is to understand how these different pieces work together to facilitate protein sorting, vesicle formation and virus budding. ■

Steven L. Alam and Wesley I. Sundquist are in the Department of Biochemistry, University of Utah, 15 North Medical Drive East, Room 4100, Salt Lake City, Utah 84112-5650, USA.
e-mails: alam@biochem.utah.edu; wes@biochem.utah.edu

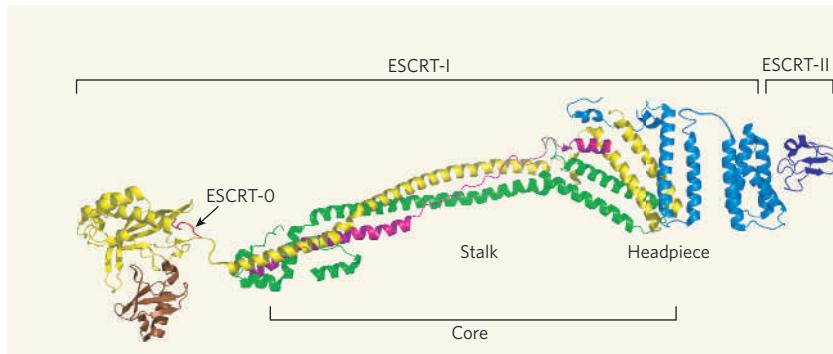


Figure 2 | Structure of the yeast ESCRT-I complex. Kostelansky *et al.*¹ found that the ESCRT-I core consists of two regions: a headpiece and a stalk. Previous studies have revealed the structures of the two independent protein-binding domains that are connected to the core through flexible tethers. The Vps23 subunit (yellow) of ESCRT-I binds to ubiquitin (brown) and to a sequence motif (red) present in the Vps27 subunit of the upstream ESCRT-0 and in p6^{Gag} proteins of HIV. The carboxy terminus of its Vps28 subunit (blue) interacts with a motif (purple) in the Vps36 subunit of the downstream ESCRT-II.

1. Kostelansky, M. S. *et al.* *Cell* **129**, 485–498 (2007).
2. Kirchhausen, T. *Nature Rev. Mol. Cell Biol.* **1**, 187–198 (2000).
3. Morita, E. & Sundquist, W. I. *Annu. Rev. Cell Dev. Biol.* **20**, 395–425 (2004).
4. Hurley, J. H. & Emr, S. D. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 277–298 (2006).
5. Williams, R. L. & Urbé, S. *Nature Rev. Mol. Cell Biol.* **8**, 355–368 (2007).
6. Morita, E. *et al.* *Cell Host Microbe* (in the press).
7. Kostelansky, M. S. *et al.* *Cell* **125**, 113–126 (2006).
8. Teo, H. *et al.* *Cell* **125**, 99–111 (2006).
9. Gill, D. J. *et al.* *EMBO J.* **26**, 600–612 (2007).
10. Oestreich, A. J. *et al.* *Mol. Biol. Cell* **18**, 646–657 (2007).
11. Curtiss, M., Jones, C. & Babst, M. *Mol. Biol. Cell* **18**, 636–645 (2007).
12. Chu, T., Sun, J., Saksena, S. & Emr, S. D. *J. Cell Biol.* **175**, 815–823 (2006).

**Cover illustration**

The integrity of inheritance, like the strength of a spider web, is weakened by disruption or damage.
(Courtesy of J. Young and A. Nussenzweig.)

Editor, *Nature*

Philip Campbell

Insights Publisher

Sarah Greaves

Publishing Assistant

Claudia Banks

Insights Editor

Ritu Dhand

Production Editors

Davina Dadley-Moore

Sarah Archibald

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Sponsorship

Gerard Preston

Emma Green

Production

Susan Gray

Marketing

Katy Dunningham

Elena Woodstock

Editorial Assistants

Jayne Hill

Alison McGill

DNA REPLICATION AND REPAIR

In the beginning, there was nucleic acid. Evolutionary biologists have argued persuasively that nucleic acids were the original self-replicating entity, an idea encapsulated in the RNA-world hypothesis. And it remains true to this day that the genome is the fundamental component required for the existence of a cell or a virus.

Given this heavy responsibility for the maintenance of life, ensuring the integrity of the genome from one generation to the next (the occasional beneficial mutation notwithstanding) is crucial. There are two aspects to this process: accurate replication of the nucleic acid so that daughter cells inherit a complete genome, and reversal of any damage that might hinder replication or alter the encoded information.

The reviews in this Insight cover aspects of repair and replication that contribute to accurate propagation of genetic information. As we have learned in the past decade, these processes are highly interconnected.

We begin at the end — at the chromosome ends, or telomeres, where the cell uses special ‘tricks’ during replication, to prevent the shortening of chromosomes. In addition, the cell needs to ignore the similarity between the ends of the telomeres and double-strand breaks in DNA so that the telomere ends are not ‘repaired’ inappropriately. The genome can also be disrupted during replication of the rest of the chromosome, through increased instability of triplet repeats, and several mechanisms suppress this instability. In addition to such endogenous disruption, DNA can be damaged by factors such as oxygen radicals or ultraviolet radiation. DNA lesions caused by such exogenous factors need to be repaired so that they do not impede the replication machinery. Finally, in the larger context, cellular DNA is packaged in a proteinaceous sheath, so interesting questions arise about how the replication and repair complexes gain access to the DNA.

When replication goes awry or DNA lesions occur, ageing and disease can result. Studies in this field might help us to mitigate such outcomes in the future.

Angela K. Eggleston, Senior Editor

REVIEWS

924 Replication and protection of telomeres

R. E. Verdun & J. Karlseder

932 Expandable DNA repeats and human disease

S. M. Mirkin

941 Base-excision repair of oxidative DNA damage

S. S. David, V. L. O’Shea & S. Kundu

951 Chromatin dynamics and the preservation of genetic information

J. A. Downs, M. C. Nussenzweig & A. Nussenzweig

**nature
insight**

Replication and protection of telomeres

Ramiro E. Verdun¹ & Jan Karlseder¹

During the evolution of linear genomes, it became essential to protect the natural chromosome ends to prevent triggering of the DNA-damage repair machinery and enzymatic attack. Telomeres — tightly regulated complexes consisting of repetitive G-rich DNA and specialized proteins — accomplish this task. Telomeres not only conceal linear chromosome ends from detection and inappropriate repair but also provide a buffer to counteract replication-associated shortening. Lessons from many model organisms have taught us about the complications of maintaining these specialized structures. Here, we discuss how telomeres interact and cooperate with the DNA replication and DNA-damage repair machineries.

The protection of genetic information is essential for cells and organisms, because the accumulation of chromosomal aberrations leads to genomic instability. The evolution of linear chromosomes presents a complex puzzle, because chromosome ends need to be protected from enzymatic attack to avoid the loss of genetic information. In addition, all cells have developed mechanisms to detect DNA lesions, and natural chromosome ends need to be hidden from this machinery. The solution in most eukaryotes lies in nucleoprotein complexes known as telomeres, which consist of G-rich DNA repeats covered by specialized binding proteins. The actual terminus of a telomere is not blunt-ended but consists of a single-stranded 3' protrusion of the G-rich strand (or G strand), known as a G tail or G overhang. These overhangs have been observed in humans, mice, ciliates, yeast, trypanosomes and plants, demonstrating that they are evolutionarily conserved and an essential feature of telomeres.

Loss of telomere function has various consequences in many model organisms, such as loss of the telomere G overhang, resection of the C-rich strand (or C strand), increased levels of recombination at chromosome ends, altered gene-expression patterns, fusion of chromosomes, instability of the genome, growth arrest and cell death. Most of our knowledge about telomere structure and function is derived from studies in a diverse range of organisms, such as the ciliates *Euplotes crassus* (also known as *Moneuplotes crassus*), *Tetrahymena thermophila* and *Oxytricha nova* (also known as *Sterkiella nova*), and the yeasts *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Kluyveromyces lactis*. In human and mouse cells, the only direct consequences of telomere dysfunction that have been identified so far are degradation of the G strand and/or chromosome fusion. Whether similar phenotypes to those observed in the ciliates and/or yeasts also occur in mammalian cells with impaired telomere function remains to be seen.

Studies in human and mouse cells suggest that the G-rich single-stranded telomere overhang can invade homologous double-stranded telomeric tracts, resulting in a large lasso-like structure, known as a telomeric loop (t-loop)¹ (Fig. 1). This provides an elegant and appealing mechanism by which chromosome ends could be protected. However, at present, it is not clear whether t-loops are present at all chromosome ends, whether they are required for chromosome protection, or whether, instead, they have a role in regulating other features of telomeres (for example, access for the telomere-specific reverse transcriptase, known as telomerase, and therefore telomere length). Telomeres in cells from humans, mice, ciliates, trypanosomes and plants, as well as yeast engineered to have long telomeres, have been shown to have t-loops².

Terminal loops are an attractive model for a specialized configuration of chromosome ends; however, it is clear from the ciliate *O. nova* that other equally efficient structures have evolved. *O. nova* chromosome ends are tightly bound by a complex of two proteins of 56 and 41 kDa, efficiently protecting both the single-stranded telomere overhang and the double-stranded telomeric DNA from modifying enzymes^{3–6}. This suggests that t-loops have evolved as only one of several means of chromosome end protection.

Mammalian telomeres are associated with the shelterin complex, a complex of interdependent telomeric core proteins consisting of telomeric-repeat-binding factor 1 (TRF1), TRF2, TRF1-interacting protein 2 (TIN2), the transcriptional repressor/activator protein RAP1, protection of telomeres 1 (POT1) and the POT1- and TIN2-organizing protein TPP1 (ref. 7; Fig. 1). TRF1 was originally reported to be involved mainly in the control of telomere length, and TRF2 was mainly implicated in chromosome end protection, by preventing end-to-end fusions^{8,9}. Now, however, these distinctions are less clear-cut, because TRF2 has been shown to have a role in telomere length regulation¹⁰, and the targeted deletion of *Trf1* in mice leads to early embryonic lethality. This lethality is probably due to telomere deprotection, because concomitant deletion of the DNA-damage sensor p53 extended the life of the embryos¹¹. Such overlapping phenotypes can be explained by the nature of the shelterin complex, which is destabilized by the removal of individual members.

In addition to shelterin, mammalian telomeres interact with a number of other factors that can influence chromosome end integrity and dynamics, such as tankyrase 1 and tankyrase 2, poly(ADP-ribose) polymerase (PARP), meiotic recombination 11 homologue (MRE11), the RecQ-like helicases WRN (Werner's syndrome protein) and BLM (Bloom's syndrome protein), Ku70, Ku86, DNA-dependent protein kinase (DNA-PK; also known as PRKDC), ataxia-telangiectasia mutated (ATM), ATM and Rad3-related (ATR), excision repair cross-complementing 1 (ERCC1), RNA-polymerase σ⁷⁰ factor (XPF) and the DNA-repair protein RAD51D⁷ (Fig. 1). This plethora of factors, many of which are involved in DNA recombination and repair, not only demonstrates the flexibility and dynamic nature of the complex but also presents a paradox, because telomeres have long been defined as structures that are protected against becoming substrates for DNA repair or recombination. However, it is becoming increasingly clear that the repair and recombination machineries are an important component of telomere replication, protection and stability. One challenge for the telomere field is to address how these machineries contribute to these different classes of DNA end.

¹The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037-1099, USA.

The end-replication problem and ageing

In 1972, James Watson wrote, "While 5' to 3' oriented growth should proceed smoothly to the end of its template, I see no simple way for 3' to 5' growth to reach the 3' end of its template"¹². Thus, he correctly predicted that the lagging strand of linear chromosomes copied by the semi-conservative replication machinery would not be fully replicated¹². In 1973, A. M. Olovnikov proposed the 'marginotomy theory of ageing', suggesting that 'telogenes' located at opposite ends of DNA molecules carry no genetic information and fulfil a buffer function. He stated that these telogenes are stochastically shortened during each mitotic cycle, providing a mechanism for ageing¹³.

Observations made by L. Hayflick in 1961 suggested that human cells derived from embryonic tissues can only divide about 50 times, and this became known as the Hayflick limit¹⁴. Since then, the assumption that the Hayflick limit is determined by the initial length of the telomeres and the rate of telomere shortening, as laid out in the mathematical approach of A. M. Olovnikov¹³, has been proved experimentally^{15,16}. It is well established that critically short telomeres cease to function as protective units and cause the cell to die or to arrest permanently.

Telomeres are now known to have many more roles than simply buffering against DNA loss; however, the initial concept of replication-associated telomere shortening was correct. At present, we refer to the inability of conventional DNA polymerases to replicate linear molecules fully as the 'end-replication problem'. This is caused by the deletion of the RNA primer of the most distal Okazaki fragment and results in the loss of about five bases of terminal genetic material per population doubling^{15,17–19}. However, the sequence loss that is predicted to occur as a result of the end-replication problem is considerably less than that which has been observed in primary human cells, which lose about 100–200 bases of TTAGGG repeats per cell division^{20–22}. Consequently, replication-associated terminal sequence loss is caused by a combination of the end-replication problem and the processing that must occur to create the G overhang on the telomeres generated by leading- and lagging-strand synthesis.

When telomeres become critically short, they are detected by the cellular DNA-damage repair machinery²³. As demonstrated in *S. cerevisiae*, chromosomes that lose a telomere are often eliminated, despite checkpoint and DNA-damage repair machineries²⁴. In human cells, p53- and RB1 (retinoblastoma 1)-dependent pathways are responsible for monitoring telomere function, whereas p53 seems to be the main sensor in mouse cells²⁵. The minimal functional telomere length, and whether this length varies among cell types, has not been clearly defined. But even in senescent human cells, telomeric double-stranded repeats are readily detectable, suggesting that several kilobases of TTAGGG repeats are required at all times. Similarly, when the telomere-protection factor TRF2 is overexpressed in telomerase-negative primary human fibroblasts, telomere-shortening rates almost double, and cells enter senescence

with considerably shorter telomeres than control populations, indicating that telomere structure, not telomere length, is the main determinant of functional telomeres²⁶.

Analysis of signalling from experimentally induced dysfunctional mammalian telomeres and from chromosome ends in senescent cells suggests that the same machinery, the intracellular DNA-damage-monitoring system, recognizes both. Telomeres that are stripped of the protective shelterin complex by expression of a dominant form of TRF2 become associated with factors involved in DNA-damage responses such as the p53-binding protein 53BP1, the histone protein γ-H2AX (phosphorylated H2AX) (see page 951), RAD17, ATM and MRE11, and are visualized as TIF, or telomere-dysfunction-induced foci²⁷. Inhibition of the phosphatidylinositol-3-OH-kinase-like kinases ATM and ATR reduces TIF formation, confirming that dysfunctional telomeres are detected by the ATM-p53 pathway²⁷. Cells carrying senescent telomeres trigger a cellular response remarkably similar to that elicited by double-strand DNA breaks²⁸; in these cells, several DNA-damage-response factors congregate at the eroded telomeres similarly to TIF. These findings underscore the importance of the DNA-damage repair machinery for telomere function, emphasizing the fact that this machinery carries out several essential tasks.

Relicative senescence can be viewed as a mechanism to limit the potential number of population doublings a cell can undergo, hypothetically rendering it a powerful tumour-suppressor mechanism²⁹. Each time a cell divides, telomeres shorten as a result of the end-replication problem and end processing. After telomeres have become critically short, they are detected by the DNA-damage repair machinery, and the cell dies or enters senescence. At present, senescence in human cells is regarded as an irreversibly arrested state, effectively inhibiting the generation of immortal cells and therefore cancer formation. As a result, major tumour-suppressive mechanisms need to be deactivated before a cell can overcome this block to immortality. Cells that continue to divide past their normal replicative limit lose all remaining protective telomeric DNA and enter a stage termed crisis. Crisis is marked by massive genomic instability and cell death. Eventually, transformed clones emerge, and although the activation of telomerase is not essential for the acquisition of a transformed phenotype³⁰, most cells that successfully exit from crisis have upregulated its activity. This observation emphasizes the dual role of telomerase in the immortalization process. On the one hand, reactivation of the enzyme in cells with critically short telomeres allows genetically unstable and immortal clones to be established, which is a major step towards cancer. On the other hand, switching on telomerase in cells that have not reached crisis prevents telomere-mediated genomic instability, which is a hallmark of cancer cells³¹. Consequently, it could be argued that telomerase fulfills a tumour-suppressive role before it contributes to the establishment of immortality (Fig. 2).

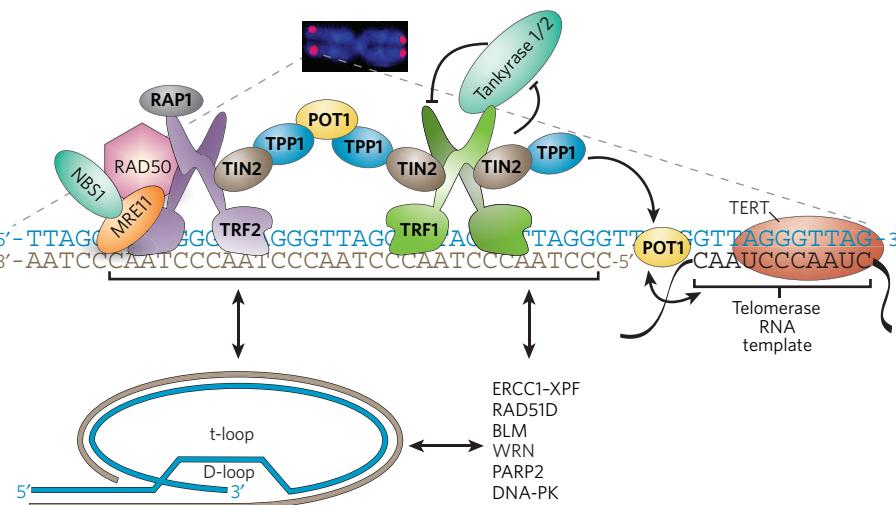


Figure 1 | The mammalian telomeric complex. The fluorescence image shows the location of a telomere within a chromosome. Mammalian telomeres consist of TTAGGG repeats with a single-stranded 3' overhang of the G-rich strand. Specific protein complexes bind to the double- and single-stranded telomeric DNA. The components of the shelterin complex are shown in bold text. The single-stranded overhang can invade the double-stranded portion of the telomere, forming protective loops — such as t-loops with displacement loops (D-loops) — at the invasion site. The telomerase complex (which contains the telomerase RNA template and the reverse transcriptase TERT) interacts with the overhang and is regulated by shelterin and other telomeric proteins⁷. Other factors that can interact with telomeres are listed. Bidirectional arrows indicate interactions.

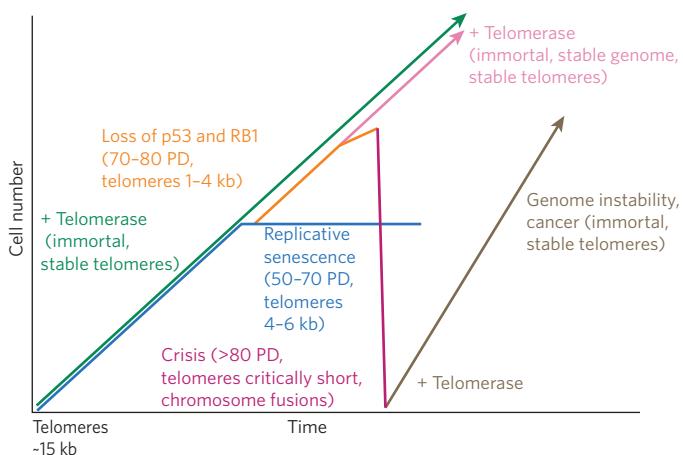


Figure 2 | Telomere shortening, senescence and cancer. Primary cells divide exponentially, and telomeres shorten from ~15 kilobases (kb) until they reach a critical length, 4–6 kb. Irreversible cell-cycle arrest then occurs (blue). Activation of telomerase before senescence allows cells to divide indefinitely and maintain a stable genome (green). If, instead, the p53 and RB1 pathways are suppressed, cells continue dividing (orange) until end protection is completely lost, resulting in telomeric crisis, cell death and massive genomic instability (dark pink). If telomerase is activated before erosion is complete, this rescues the genome from instability by re-establishing telomere maintenance (light pink). Activation of telomerase after the accumulation of mutations results in an unstable genome, allowing clones that carry multiple mutations to escape cell death (that is, to become immortal). Such cells are predisposed to oncogenic transformation (brown). PD, population doublings.

Approximately 10% of human tumours rely on a telomerase-independent method to maintain their telomeres. Known as ALT (alternative lengthening of telomeres), it is based on recombination between telomeres³² (Box 1). Tumours resulting from *Wrn*^{-/-} mouse cells without telomerase activity readily engage the ALT pathway, and although the exact molecular mechanism is not understood, preliminary findings point to aberrant homologous recombination as the underlying cause^{33,34}.

Despite the finding that expression of the catalytic subunit of telomerase readily immortalizes primary cells without seeming to cause genomic instability^{16,35}, the *in vivo* evidence for telomere involvement in human ageing is limited to correlations. For example, the mean length of telomeric restriction fragments from DNA isolated from sperm cells is considerably longer than comparable fragments isolated from replicating cells *in vivo*³⁶, and it has recently been demonstrated that senescent cells account for up to 15% of the cell population in the skin of aged baboons³⁷. Many such correlations have been documented, and although they show that telomeres shorten with age, it is unclear whether telomere shortening causes ageing *in vivo*.

Little is known about the role of telomeres during the ageing process of differentiated cells and in organisms that do not contain mitotic cells. Although it has been suggested that telomere elongation extends the lifespan of the nematode *Caenorhabditis elegans*³⁸ (in which cells do not undergo mitosis after development is complete), this is probably a secondary effect of *hrp-1* overexpression in these animals, because clonal wild-type nematode strains with varying telomere length did not show any differences in organismal ageing and lifespan³⁹. Consequently, and in line with Watson's and Olovnikov's concepts of the end-replication problem, it is unlikely that telomeres have a major role in the ageing of non-dividing cells.

Telomere replication

The end-replication problem correctly predicts that linear DNA molecules shorten during every replicative cycle. Consequently, in the absence of a mechanism to maintain the absolute ends, chromosomes eventually lose the protective cap provided by their ends, resulting

in the loss of genomic integrity. To counteract replication-associated telomere shortening, telomerase evolved. Telomerase is a specialized reverse transcriptase complex and can add G-rich telomeric repeats to the absolute ends of chromosomes using its own internal RNA template, effectively stabilizing telomere length. In *S. cerevisiae*, telomeres switch between extendable and non-extendable states⁴⁰. In human cells, the limited amount of telomerase contributes to telomere length homeostasis, because increased amounts of telomerase change telomere length settings to a different equilibrium⁴¹. In *S. cerevisiae*, deletion of the gene encoding the telomerase RNA template (*TLC1*) or the catalytic subunit (*EST2*) leads to gradual telomere shortening and growth arrest^{17–19}. Similarly, mice lacking the gene encoding either the telomerase RNA (*Terc*) or the reverse transcriptase domain (*Tert*) gradually lose their telomeres over several generations, resulting in degeneration of highly proliferative cell populations and sterility^{42–44}. Expression of human *TERT* in fibroblasts causes telomere elongation and renders the cells immortal, effectively avoiding telomere-shortening-dependent replicative senescence¹⁶.

No origin of replication has been detected in telomeres, rendering the closest origin, resident in the subtelomeric region of the chromosome, the starting point for the replication of chromosome ends. Passage of the replication fork through the telomere is thought to generate a blunt-ended leading-strand product and a lagging-strand product with a short 3' G overhang (Fig. 3). In the presence of telomerase, and during new telomere synthesis, the actions of the conventional replication machinery and telomerase are closely coordinated. Inhibition of C-strand synthesis in the ciliate *E. crassus* by using aphidicolin, a specific inhibitor of DNA polymerase- α and DNA polymerase- δ , leads to a general lengthening of the G strand, thereby showing that C- and G-strand synthesis is coordinated⁴⁵. Similarly, addition of new telomeres by telomerase in *S. cerevisiae* requires not only extension of the 3' G-rich end by telomerase but also fill-in synthesis of the C strand by DNA primase, DNA polymerase- α and DNA polymerase- δ . G-strand polymerization by telomerase is inhibited in *S. cerevisiae* if DNA polymerase- α and DNA polymerase- δ are inactive, suggesting that telomerase needs to interact with the lagging-strand synthesis machinery to be active⁴⁶. An excellent candidate for regulating the coordination between telomerase and the conventional DNA polymerases is the Cdc13 complex, which attracts telomerase to chromosome ends in *S. cerevisiae*. Cdc13 binds to single-stranded G-rich telomeric DNA and then recruits telomerase^{47,48}. Several studies have suggested that, subsequently, lagging-strand synthesis fills in the C strand, and then inhibits telomerase in a Cdc13-dependent manner^{49,50}.

Little is known about coordinated C- and G-strand synthesis in mammalian cells, but activation of a temperature-sensitive allele encoding DNA polymerase- α causes elongation of the G tail and of the telomere overall, suggesting that coordination of telomerase with the replication machinery is a common feature in all organisms⁵¹.

The G-rich and repetitive nature of telomeric DNA complicates replication as well, because it potentially allows the formation of secondary structures, such as G quartets⁵². Consequently, telomeric proteins support the progressing replication fork, allowing efficient telomere synthesis. In *S. pombe*, Taz1, the homologue of TRF1 and TRF2, is required for telomere replication. Without Taz1, replication forks stall at telomeric sequences, regardless of whether the repeats are located at the ends or in the interior of chromosomes⁵³. This suggests that the Taz1-dependent telomere-replication phenotype is due to characteristics of the telomeric sequence itself and not to its position on the chromosome. In human cells, the RecQ-like helicase WRN contributes to efficient telomere replication. Overexpression of a helicase-defective *WRN* allele causes the occasional loss of telomeres generated by the lagging-strand machinery, and telomerase expression compensates for this loss of telomeric sequence, implicating telomere maintenance in the pathology of Werner's syndrome⁵⁴. Accordingly, targeted deletion of *Wrn* in mice leads to phenotypes that resemble the human Werner's syndrome only when telomerase is also deleted⁵⁵. In summary, it is becoming increasingly clear that telomere replication and telomerase-dependent telomere elongation are highly coordinated processes and that telomeric proteins have essential roles in the regulation of these processes.

The generation of telomere overhangs

After telomere replication is completed, the newly generated lagging-strand telomere carries a short 3' overhang, resulting from the removal of the most distal RNA primer used for Okazaki fragment synthesis. It is not clear whether this distal primer is placed at the absolute terminus of the chromosome or a few bases from the end, therefore allowing an overhang that could be longer than the length of the RNA fragment. By contrast, leading-strand synthesis is expected to continue until it reaches the end of the template, resulting in blunt-ended products (Fig. 3). In ciliate, yeast and human cells, overhangs can be detected at both ends of the chromosomes, suggesting that there are regulated mechanisms for G-tail generation^{56–60}. Currently, no candidates for overhang-generating nucleases have been identified in any organism, but the field is looking to ciliates for clues, because much of the work on telomere replication was pioneered in these organisms.

T. thermophila maintains G overhangs with defined sequence and length on both chromosome ends⁵⁶. Because both the G strand and the C strand are processed accurately in the absence of telomerase, it has been suggested that C-strand resection works in collaboration with G-strand cleavage to generate a functional telomere end. These steps potentially include more than one nuclease, and when the telomeric sequence is artificially changed, the terminal nucleotides were not altered, suggesting that the nuclease activities do not show sequence preference and that additional factors, such as proteins that bind to the single-stranded overhang, regulate specificity⁶¹.

In human cells, in the absence of telomerase, the leading-strand daughter telomeres carry longer overhangs than the telomeres synthesized by the lagging-strand machinery⁶². However, when the catalytic subunit of telomerase is introduced, similar lengths are found at both leading and lagging daughter telomeres. Moreover, both daughter telomeres have conserved terminal nucleotides at their 5' ends. Small-interfering-RNA-dependent knockdown of expression of the shelterin component POT1 in human tumour cells randomizes the last nucleotides of the 5' telomere end, suggesting that this single-stranded TTAGGG-binding protein is involved in regulation of terminal specificity⁶³. This process differs for the 3' end, where the terminal residues seem to be much more variable in the absence of telomerase. Overexpression or suppression of individual components of the telomere-binding protein complex or of the multisubunit telomerase complex might already disturb the equilibrium at telomeres, limiting the conclusions that can be drawn about the *in vivo* situation from such experiments.

Eventually, detailed knowledge about overhang length and base specificity of leading- and lagging-strand telomeres in the presence and absence of telomerase will provide insight into the coordination of telomerase activity at telomeres. There are several possible ways in which this coordination might occur. Both daughter telomeres are nucleolytically recessed in the 5' to 3' direction after replication, but the efficiency varies at each strand, potentially rendering the leading-strand telomere with a longer overhang⁶². The presence of telomerase complicates processing, because telomerase could take advantage of the short overhang at the lagging strand before nuclease action, specifically elongating telomeres replicated by the lagging-strand machinery. However, if telomerase acts after overhangs are generated on both strands, more than one fill-in step by the lagging-strand machinery and more than one resection step might be required to generate functional chromosome ends. The field of telomere research is in agreement that well-regulated overhang generation is an essential step for telomere function and thus for chromosome protection; however, at this stage, only the surface of this complex problem has been scratched.

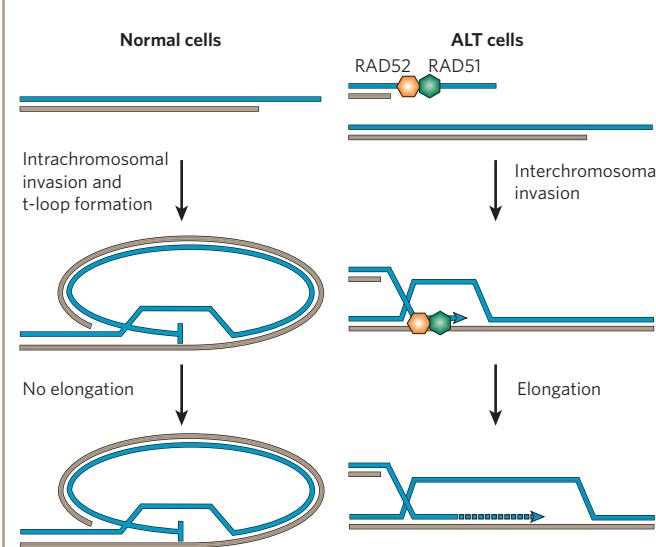
Telomeres and the DNA-damage repair machinery

The protective features of telomeres are lost when chromosome ends become uncapped, through the mechanisms described here. These subsequently dysfunctional telomeres are then subject to DNA repair by non-homologous end joining (NHEJ) or homologous recombination.

In accordance with a requirement for the NHEJ pathway (which depends on Ku, DNA-PKcs (the catalytic subunit of DNA-PK), and DNA

ligase IV and its cofactor XRCC4) for the processing of dysfunctional telomeres, DNA ligase IV is required for end-to-end fusion of critically short or dysfunctional telomeres. DNA ligase IV fuses telomeres in *S. pombe* lacking *taz1* (a homologue of the mammalian *TRF* genes)⁶⁴, in *S. cerevisiae* with mutated *TEL1* (which encodes a protein kinase) or *MEC1* (which encodes a signal transducer)⁶⁵, and in mouse cells that lack *TRF2* and therefore contain uncapped chromosome ends⁶⁶. Ku contributes to telomere protection in *S. cerevisiae*, *S. pombe* and mammalian

Box 1 | ALT, homologous recombination and BIR at telomeres



A subset of immortalized cells do not show telomerase upregulation and use a recombination-based pathway known as alternative lengthening of telomeres (ALT) to maintain chromosome ends. One of the main characteristics of 'ALT cells' is the presence of promyelocytic leukaemia (PML) bodies, which are subnuclear structures that contain telomeric DNA, telomeric proteins and factors involved in DNA recombination and repair. ALT cells have heterogeneous telomere lengths, ranging from critically short telomeres to telomeres of up to 100 kb, as well as telomeric DNA circles of 1–60 kb⁹⁴. It has been suggested that the homologous-recombination machinery is responsible for the amplification of the telomeric sequences in ALT cells, because a selection marker introduced into telomeres of ALT and non-ALT cells spreads throughout telomeres only in ALT cells³³. ALT cells show an increased rate of sister chromatid exchange^{95,96}, suggesting that the homologous-recombination pathway is involved. Overexpression of a mutant *TRF2* allele in non-ALT cells generates t-loop-sized DNA circles that depend on the ERCC1-XRCC3 complex⁸¹. This phenotype resembles the ALT-associated DNA circles and again suggests a possible role for homologous recombination in the ALT pathway.

The molecular mechanism of the ALT pathway is far from understood but resembles break-induced replication (BIR). BIR is a gene-conversion mechanism that is induced only when one DNA end invades a homologous sequence and initiates DNA replication with the homologous sequence as template. Proteins such as RAD50, RAD52 and MRE11 are involved in BIR⁹⁷. It is feasible that invasion of the single-stranded telomere overhang into double-stranded TTAGGG repeats is a BIR-like situation. Consequently, it is possible that the telomere maintenance mechanism used by ALT cells is similar to that proposed for BIR. The figure compares telomere structure and regulation in non-ALT cells and ALT cells.

In the absence of telomerase, a subset of yeast cells elongate their telomeres through amplification of telomeric and subtelomeric repeat sequences. This recombination maintenance pathway depends on RAD52 (ref. 98), suggesting that homologous recombination can elongate telomeres in a telomerase-independent manner.

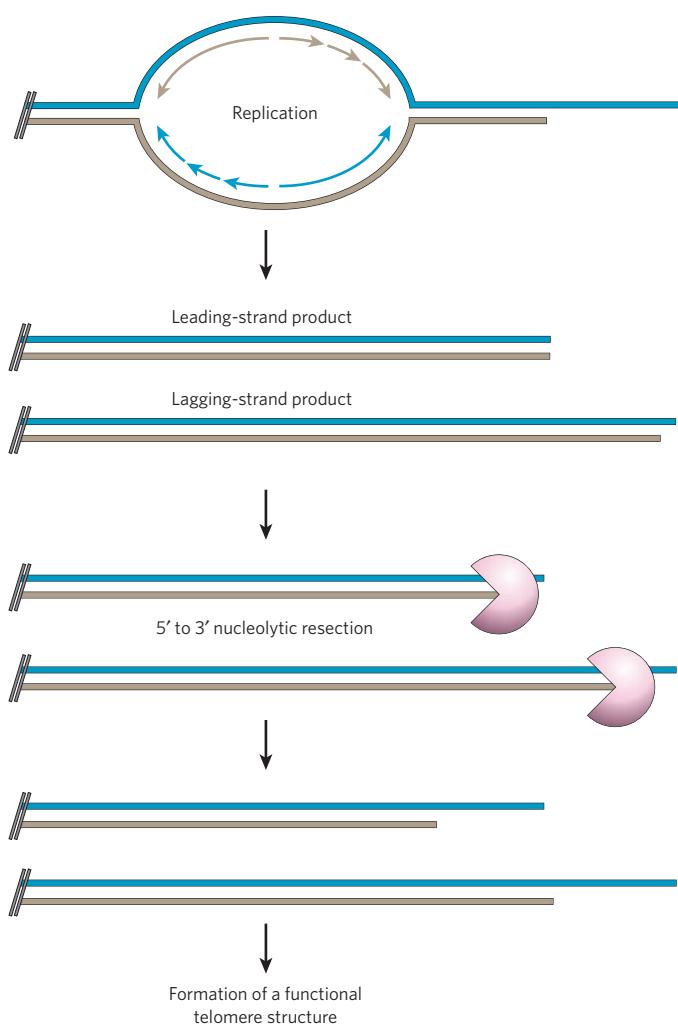


Figure 3 | End replication and processing. There are 3' G overhangs at both ends of the chromosome, and these are thought to be generated by 5' to 3' nucleolytic activity. Semi-conservative replication of telomeres generates a blunt-ended leading-strand product and a lagging-strand product with a short overhang. Nucleolytic digestion (pink) in the 5' to 3' direction then generates G overhangs, which allow the formation of a functional telomere structure (not shown). The short overhang generated by lagging-strand synthesis could be sufficient for a functional telomere, so it has been proposed that only the leading-strand product undergoes nucleolytic digestion.

cells. In all of these organisms, Ku associates with telomeric DNA, regulates telomerase and inhibits inappropriate recombination and repair events at telomeres^{67,68}.

In *S. cerevisiae*, the single-stranded telomeric DNA-binding factor Cdc13, which also has a prominent role in telomerase recruitment, is central to chromosome end protection⁶⁹. A protective complex of Cdc13, Stn1 and Ten1 protects the telomere from excessive degradation of the C strand^{47,70,71}, resulting in long G tails, which trigger Rad9-dependent cell-cycle arrest^{72,73}.

Loss of the *S. pombe* double-stranded telomere-binding protein Taz1 results in uncontrolled elongation of single- and double-stranded telomeric tracts⁷⁴, as well as loss of viability and telomere fusion in yeast arrested in the G1 phase of the cell cycle⁶⁴. Chromosome end fusions are also a consequence of loss of Pot1, a single-stranded telomere-binding protein, which is a homologue of the α -subunit of the *O. nova* end-binding factor⁷⁵. *S. pombe* cells lacking Pot1 rapidly lose telomeric and subtelomeric DNA and become inviable, although a subset of cells survive and have circularized chromosomes. Human POT1 regulates telomere length in a telomerase-dependent manner, relaying information

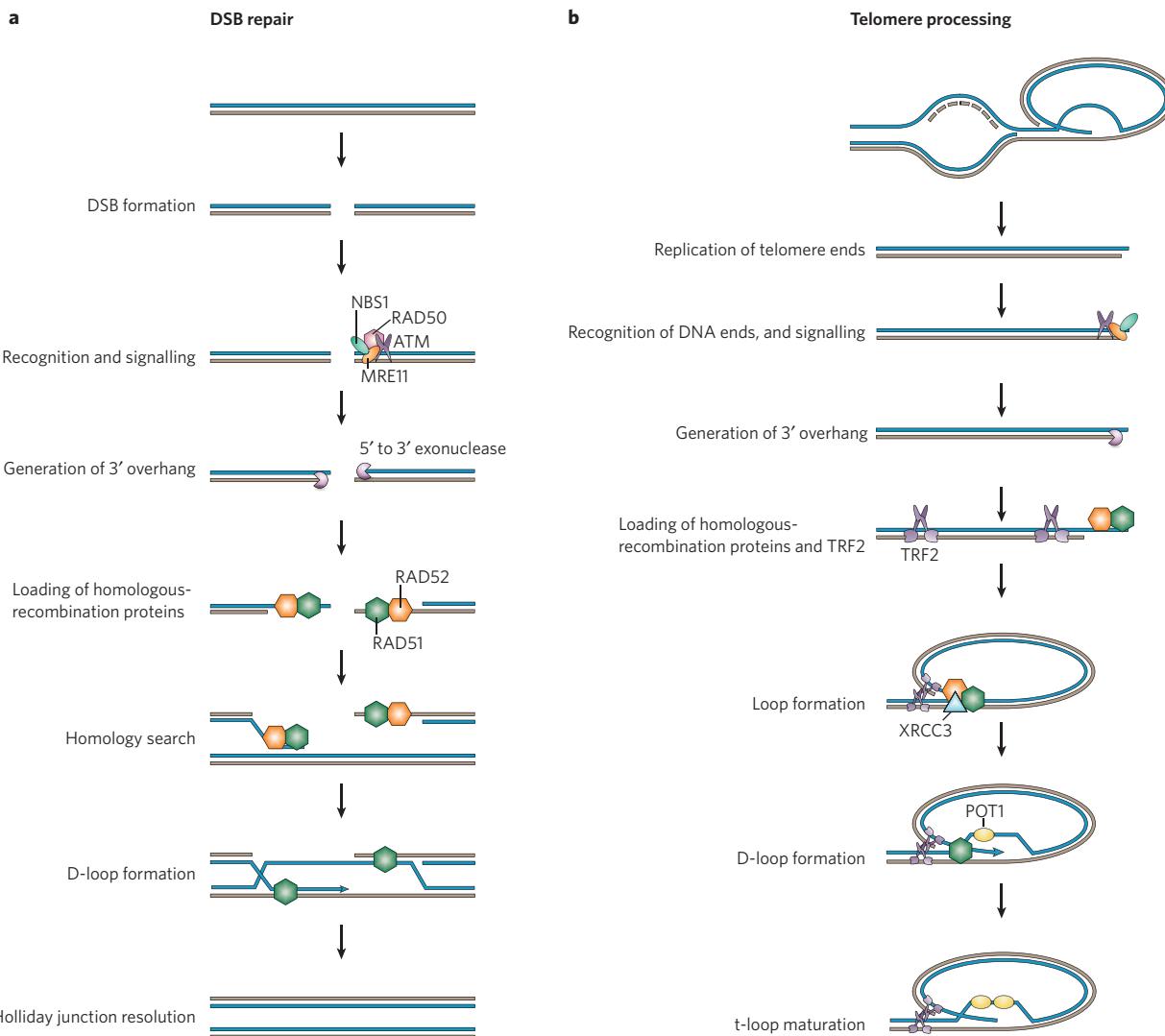
from shelterin to the telomerase complex^{75–77}. Mouse cells have two POT1 proteins, POT1A and POT1B, suggesting recent expansion of the telomeric complex in rodents^{78,79}. POT1A is sufficient to repress DNA-damage signalling at telomeres in the absence of POT1B. However, lack of POT1B leads to a substantial increase in the length of the G tail in a telomerase-independent manner, a phenotype that could not be rescued by overexpression of POT1A, establishing that the two proteins have distinct roles in telomere protection. In addition to POT1, the mammalian factor TRF2 has a substantial role in protecting telomeres from NHEJ and homologous recombination. Telomeres in mouse and human cells that lack TRF2 lose the G tail, are detected as damage sites, and are substrates for DNA-ligase-IV-dependent fusion^{9,66,80}. However, telomeres in cells that lack both DNA ligase IV and TRF2 do not show overhang loss, although they are still recognized as DNA damage, indicating that the loss of the G tail is a consequence of NHEJ and is not required for the DNA-damage pathways at telomeres⁸⁰. In addition, overexpression of a human TRF2 allele encoding a protein that lacks the amino terminus but retains the DNA-binding region causes the rapid shortening of telomeres and the generation of extrachromosomal telomeric DNA circles⁸¹. These circles can readily be detected in cells in which the ALT pathway is engaged (Box 1) and depend on the homologous-recombination protein XRCC3, which is involved in the resolution of Holliday junctions (junctions between four DNA strands), suggesting that TRF2 protects telomeres not only from NHEJ but also from homologous-recombination-based deletion of large stretches of DNA.

It is now well established that the natural chromosome ends need to be protected from inappropriate repair, so it seems paradoxical that several proteins involved in the detection or repair of DNA lesions localize to telomeres. Such localizing proteins include the MRX/N complex (Mre11–Rad50–Xrn2 in yeast and MRE11–RAD50–NBS1 in mammals) and the protein kinases Tel1 and Mec1 in *S. cerevisiae* or their mammalian homologues, ATM and ATR, all of which have crucial roles in DNA-damage signalling and telomeric integrity. Deletion of *MRE11* or *RAD50* in *S. cerevisiae* leads to gradual telomere shortening, which does not increase when *EST2* is also deleted, placing the MRX complex in the same pathway as telomerase⁸², probably recruiting the enzyme to telomeres⁸³. The finding that MRX localizes to telomeres in late S phase suggests that this complex is required to prepare the telomere for telomerase-dependent replication⁸⁴. Co-deletion of MRN components and *rad3* (the homologue of *MEC1* and *ATR*) in *S. pombe* results in telomere loss and chromosome circularization⁸⁵, again placing the complex in a pathway that provides protection against telomere loss. Similarly, Tel1 and Mec1 can be found at *S. cerevisiae* telomeres⁸⁶, and deletion of the genes encoding both molecules causes telomere loss, a phenotype that can be overcome by attracting active telomerase to the chromosome ends⁸⁷. At the same time, Tel1 protects telomeres against NHEJ-dependent fusion, because the frequency of telomeres fused to an inducible double-strand break increases sharply in a strain that lacks both Tel1 and Tlc1 (ref. 88). In humans, cells derived from patients with cancer-prone syndromes such as the Nijmegen breakage syndrome or ataxia-telangiectasia (which carry mutations in *NBS1* and *ATM*, respectively) show accelerated telomere shortening and chromosome fusions^{89,90}. Similarly to yeast, the MRN complex and ATM have been detected at telomeres^{91,92}, suggesting dual roles in DNA-damage signalling and telomere maintenance for these factors.

All of these observations indicate that functional telomeres require interaction with DNA-damage repair proteins, suggesting that the DNA-damage repair machinery is involved in replication of telomeres, protection of functional chromosome ends, and detection of, and signalling from, dysfunctional ones.

Similarities between the proteins responsible for the detection and repair of DNA lesions and those found at functional and dysfunctional telomeres suggest that, for a cell, the difference between a DNA break and a telomere is less pronounced than previously assumed. The finding that functional telomeres are detected by the DNA-damage repair machinery in every cell cycle, and the presence of the homologous-recombination machinery at telomeres in G2 phase, suggests that this pathway is involved in telomere end processing^{92,93} (Box 2).

Box 2 | Telomere maintenance and double-strand-break repair



The repair of a double-strand break (DSB) by the homologous-recombination repair pathway involves well-characterized molecular steps and proteins (see figure, part **a**). After a DSB is detected, the MRN complex (MRE11-NBS1-RAD50) is one of the first repair factors to be recruited, and this is followed by MRN-dependent activation of the protein kinases ATM and ATR⁹⁹. Activation of these protein kinases results in signals that lead to the recruitment of processing factors, which generate 3' single-stranded overhangs. The exposed overhangs are coated by replication protein A, which protects the DNA against degradation and inhibits the formation of secondary structures (not shown). Next, the single-stranded DNA invades homologous duplex DNA sequences, forming a displacement loop (D-loop) by homologous pairing and strand exchange, a process catalysed by RAD51 and stimulated by RAD52, RAD54 and RAD55-RAD57. Using the homologous sequence as a template, the invading strand primes DNA synthesis, generating a Holliday junction. The genetic

information is restored after the Holliday junction is cleaved by RAD51C-XRCC3-dependent resolvase activity¹⁰⁰.

Detection of ends, generation of 3' overhangs and invasion of homologous sequences are essential steps for DSB repair, and these steps seem strikingly similar to the predicted order of post-replicative telomere processing and of t-loop formation (see figure, part **b**). Recently, we suggested that functional human telomeres in primary fibroblasts interact with proteins of the DNA-damage repair machinery⁹². MRE11, NBS1 and activated ATM localize to telomeres from late S phase until G2 phase, suggesting that telomeres are detected as DNA damage. The damage signal is localized to chromosome ends, because neither stabilization of p53 nor phosphorylation of CHK2 (checkpoint 2 homologue; also known as CHEK2) can be observed in the nucleus, suggesting that the signal is well controlled, and does not lead to cell-cycle arrest. All of these results promote the hypothesis that telomeres are not always hidden from the DNA-damage repair machinery and suggest that telomeres require a close

relationship with the DNA-damage-response pathways for function, fuelling the idea that the processing of telomeres is similar to the processing of DNA breaks.

In a manner analogous to that of DSB processing, telomeres recruit RAD51, RAD52 and XRCC3 before mitosis, potentially resulting in a search for homologous DNA sequences, followed by strand invasion. However, invasion of another chromosome by a telomere overhang can lead to a deleterious phenotype. TRF2 is a good candidate for involvement in avoidance of inter-telomere invasion, because this protein can form t-loops *in vitro* by keeping the telomere end and the duplex DNA of the same telomere in proximity. This model is supported by the finding that in an *in vitro* assay with telomeric substrates, not only is the homologous-recombination machinery required for efficient invasion but also the telomeric protein TRF2 (ref. 93). Consequently, this TRF2 activity, together with the homologous-recombination machinery, potentially facilitates the formation of the D-loop that forms the core of the t-loop structure and provides a substrate for POT1.

Conclusions

The field of telomere biology has progressed considerably from the simplistic view that telomeres function only as non-coding buffer zones at the ends of linear chromosomes. We now view telomeres as highly specialized and regulated complexes in which length and structure determine integrity and function. Despite the open questions about the requirements of t-loops for end protection or telomere length regulation, it has been accepted that the G-rich 3' single-stranded overhang is required for telomere function.

However, the cell not only needs to control the formation of the 3' overhang (and therefore the adoption of a functional telomere structure) but also needs to monitor telomere length, as demonstrated by the finding that critically short telomeres lose their protective function.

A similar bipolar relationship is observed between telomeres and DNA-damage-response pathways. On the one hand, the intracellular DNA-damage repair machinery is required to detect dysfunctional telomeres, which are consequently processed like any other double-strand break. On the other hand, the DNA-damage repair machinery is required for telomere replication and telomere protection, and it also seems to be essential for the formation of a functional telomere structure.

Taken together, all of these observations demonstrate that Watson and Olovnikov were correct when they suggested a problem in the replication of terminal DNA. The telomere, ageing and cancer fields have since managed to advance our understanding of the problem considerably, although it has not yet been solved. ■

1. Griffith, J. D. et al. Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
2. de Lange, T. T-loops and the origin of telomeres. *Nature Rev. Mol. Cell Biol.* **5**, 323–329 (2004).
3. Gottschling, D. E. & Zakian, V. A. Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell* **47**, 195–205 (1986).
4. Price, C. M. & Cech, T. R. Telomeric DNA–protein interactions of *Oxytricha* macronuclear DNA. *Genes Dev.* **1**, 783–793 (1987).
5. Gray, J. T., Celander, D. W., Price, C. M. & Cech, T. R. Cloning and expression of genes for the *Oxytricha* telomere-binding protein: specific subunit interactions in the telomeric complex. *Cell* **67**, 807–814 (1991).
6. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A. & Schultz, S. C. Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell* **95**, 963–974 (1998).
7. de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
8. van Steensel, B. & de Lange, T. Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740–743 (1997).
9. van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).
10. Smogorzewska, A. et al. Control of human telomere length by TRF1 and TRF2. *Mol. Cell. Biol.* **20**, 1659–1668 (2000).
11. Karlseder, J. et al. Targeted deletion reveals an essential function for the telomere length regulator Trf1. *Mol. Cell. Biol.* **23**, 6533–6541 (2003).
12. Watson, J. D. Origin of concatemeric T7 DNA. *Nature New Biol.* **239**, 197–201 (1972).
13. Olovnikov, A. M. A theory of marginotomy. *J. Theor. Biol.* **41**, 181–190 (1973).
14. Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp. Cell Res.* **25**, 585–621 (1961).
15. Lundblad, V. & Szostak, J. W. A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**, 633–643 (1989).
16. Bodnar, A. G. et al. Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).
17. Singer, M. S. & Gottschling, D. E. TLC1: template RNA component of *Saccharomyces cerevisiae* telomerase. *Science* **266**, 404–409 (1994).
18. Lendway, T. S., Morris, D. K., Sah, J., Balasubramanian, B. & Lundblad, V. Senescence mutants of *Saccharomyces cerevisiae* with a defect in telomere replication identify three additional EST genes. *Genetics* **144**, 1399–1412 (1996).
19. Lingner, J. et al. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561–567 (1997).
20. Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460 (1990).
21. Counter, C. M. et al. Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* **11**, 1921–1929 (1992).
22. Shay, J. W. & Wright, W. E. Hayflick, his limit, and cellular ageing. *Nature Rev. Mol. Cell Biol.* **1**, 72–76 (2000).
23. de Lange, T. Protection of mammalian telomeres. *Oncogene* **21**, 532–540 (2002).
24. Sandell, L. L. & Zakian, V. A. Loss of a yeast telomere: arrest, recovery, and chromosome loss. *Cell* **75**, 729–739 (1993).
25. Smogorzewska, A. & de Lange, T. Different telomere damage signaling pathways in human and mouse cells. *EMBO J.* **21**, 4338–4348 (2002).
26. Karlseder, J., Smogorzewska, A. & de Lange, T. Senescence induced by altered telomere state, not telomere loss. *Science* **295**, 2446–2449 (2002).
27. Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).
28. d'Adda di Fagagna, F. et al. A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).
29. de Lange, T. & Jacks, T. For better or worse? Telomerase inhibition and cancer. *Cell* **98**, 273–275 (1999).
30. Seger, Y. R. et al. Transformation of normal human cells in the absence of telomerase activation. *Cancer Cell* **2**, 401–413 (2002).
31. Artandi, S. E. & DePinho, R. A. A critical role for telomeres in suppressing and facilitating carcinogenesis. *Curr. Opin. Genet. Dev.* **10**, 39–46 (2000).
32. Reddel, R. R. & Bryan, T. M. Alternative lengthening of telomeres: dangerous road less travelled. *Lancet* **361**, 1840–1841 (2003).
33. Dunham, M. A., Neumann, A. A., Fasching, C. L. & Reddel, R. R. Telomere maintenance by recombination in human cells. *Nature Genet.* **26**, 447–450 (2000).
34. Varley, H., Pickett, H. A., Foxon, J. L., Reddel, R. R. & Royle, N. J. Molecular characterization of inter-telomere and intra-telomere mutations in human ALT cells. *Nature Genet.* **30**, 301–305 (2002).
35. Morales, C. P. et al. Absence of cancer-associated changes in human fibroblasts immortalized with telomerase. *Nature Genet.* **21**, 115–118 (1999).
36. Cooke, H. J. & Smith, B. A. Variability at the telomeres of the human X/Y pseudoautosomal region. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 213–219 (1986).
37. Herbig, U., Ferreira, M., Condel, L., Carey, D. & Sedivy, J. M. Cellular senescence in aging primates. *Science* **311**, 1257 (2006).
38. Joeng, K. S., Song, E. J., Lee, K. J. & Lee, J. Long lifespan in worms with long telomeric DNA. *Nature Genet.* **36**, 607–611 (2004).
39. Raices, M., Maruyama, H., Dillin, A. & Karlseder, J. Uncoupling of longevity and telomere length in *C. elegans*. *PLoS Genet.* **1**, e30 (2005).
40. Teixeira, M. T., Arneric, M., Sperisen, P. & Lingner, J. Telomere length homeostasis is achieved via a switch between telomerase-extensible and -nonextensible states. *Cell* **117**, 323–335 (2004).
41. Cristofari, G. & Lingner, J. Telomere length homeostasis requires that telomerase levels are limiting. *EMBO J.* **25**, 556–574 (2006).
42. Blasco, M. A. et al. Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
43. Lee, H. W. et al. Essential role of mouse telomerase in highly proliferative organs. *Nature* **392**, 569–574 (1998).
44. Liu, Y. et al. The telomerase reverse transcriptase is limiting and necessary for telomerase function *in vivo*. *Curr. Biol.* **10**, 1459–1462 (2000).
45. Fan, X. & Price, C. M. Coordinate regulation of G- and C strand length during new telomere synthesis. *Mol. Cell. Biol.* **8**, 2145–2155 (1997).
46. Diede, S. J. & Gottschling, D. E. Telomerase-mediated telomere addition *in vivo* requires DNA primase and DNA polymerases α and δ . *Cell* **99**, 723–733 (1999).
47. Pennock, E., Buckley, K. & Lundblad, V. Cdc13 delivers separate complexes to the telomere for end protection and replication. *Cell* **104**, 387–396 (2001).
48. Bianchi, A., Negrini, S. & Shore, D. Delivery of yeast telomerase to a DNA break depends on the recruitment functions of Cdc13 and Est1. *Mol. Cell* **16**, 139–146 (2004).
49. Chandra, A., Hughes, T. R., Nugent, C. I. & Lundblad, V. Cdc13 both positively and negatively regulates telomere replication. *Genes Dev.* **15**, 404–414 (2001).
50. Grossi, S., Puglisi, A., Dmitriev, P. V., Lopes, M. & Shore, D. Pol12, the B subunit of DNA polymerase α , functions in both telomere capping and length regulation. *Genes Dev.* **18**, 992–1006 (2004).
51. Nakamura, M., Nabetani, A., Mizuno, T., Hanaoka, F. & Ishikawa, F. Alterations of DNA and chromatin structures at telomeres and genetic instability in mouse cells defective in DNA polymerase α . *Mol. Cell. Biol.* **25**, 11073–11088 (2005).
52. Rhodes, D. & Giraldo, R. Telomere structure and function. *Curr. Opin. Struct. Biol.* **5**, 311–322 (1995).
53. Miller, K. M., Rog, O. & Cooper, J. P. Semi-conservative DNA replication through telomeres requires Taz1. *Nature* **440**, 824–828 (2006).
54. Crabbe, L., Verdun, R. E., Haggblom, C. I. & Karlseder, J. Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science* **306**, 1951–1953 (2004).
55. Chang, S. et al. Essential role of limiting telomeres in the pathogenesis of Werner syndrome. *Nature Genet.* **36**, 877–882 (2004).
56. Jacob, N. K., Rop, R. P. & Price, C. M. G-overhang dynamics at *Tetrahymena* telomeres. *EMBO J.* **20**, 4299–4308 (2001).
57. Wellinger, R. J., Ethier, K., Labrecque, P. & Zakian, V. A. Evidence for a new step in telomere maintenance. *Cell* **85**, 423–433 (1996).
58. Dionne, I. & Wellinger, R. J. Cell cycle-regulated generation of single-stranded G-rich DNA in the absence of telomerase. *Proc. Natl. Acad. Sci. USA* **93**, 13902–13907 (1996).
59. Makarov, V. L., Hirose, Y. & Langmore, J. P. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* **88**, 657–666 (1997).
60. Chai, W., Du, Q., Shay, J. W. & Wright, W. E. Human telomeres have different overhang sizes at leading versus lagging strands. *Mol. Cell* **21**, 427–435 (2006).
61. Jacob, N. K., Kirk, K. E. & Price, C. M. Generation of telomeric G strand overhangs involves both G and C strand cleavage. *Mol. Cell* **11**, 1021–1032 (2003).
62. Sfeir, A. J., Chai, W., Shay, J. W. & Wright, W. E. Telomere-end processing the terminal nucleotides of human chromosomes. *Mol. Cell* **18**, 131–138 (2005).
63. Hockemeyer, D., Sfeir, A. J., Shay, J. W., Wright, W. E. & de Lange, T. POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end. *EMBO J.* **24**, 2667–2678 (2005).
64. Ferreira, M. G. & Cooper, J. P. The fission yeast Taz1 protein protects chromosomes from Ku-dependent end-to-end fusions. *Mol. Cell* **7**, 55–63 (2001).
65. Mieczkowski, P. A., Mieczkowska, J. O., Dominska, M. & Petes, T. D. Genetic regulation of telomere-telomere fusions in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **100**, 10854–10859 (2003).
66. Smogorzewska, A., Karlseder, J., Holtgreve-Grez, H., Jauch, A. & de Lange, T. DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2. *Curr. Biol.* **12**, 1635 (2002).

67. Fisher, T. S., Taggart, A. K. & Zakian, V. A. Cell cycle-dependent regulation of yeast telomerase by Ku. *Nature Struct. Mol. Biol.* **11**, 1198–1205 (2004).

68. Fisher, T. S. & Zakian, V. A. Ku: a multifunctional protein involved in telomere maintenance. *DNA Repair (Amst.)* **4**, 1215–1226 (2005).

69. Lin, J. J. & Zakian, V. A. The *Saccharomyces* CDC13 protein is a single-strand TG_{n-3} telomeric DNA-binding protein *in vitro* that affects telomere behavior *in vivo*. *Proc. Natl. Acad. Sci. USA* **93**, 13760–13765 (1996).

70. Grandin, N., Reed, S. I. & Charbonneau, M. Stn1, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with Cdc13. *Genes Dev.* **11**, 512–527 (1997).

71. Grandin, N., Damon, C. & Charbonneau, M. Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13. *EMBO J.* **20**, 1173–1183 (2001).

72. Garvik, B., Carson, M. & Hartwell, L. Single-stranded DNA arising at telomeres in *cdc13* mutants may constitute a specific signal for the RAD9 checkpoint. *Mol. Cell. Biol.* **15**, 6128–6138 (1995); erratum **16**, 457 (1996).

73. Booth, C., Griffith, E., Brady, G. & Lydall, D. Quantitative amplification of single-stranded DNA (QAOS) demonstrates that *cdc13-1* mutants generate ssDNA in a telomere to centromere direction. *Nucleic Acids Res.* **29**, 4414–4422 (2001).

74. Cooper, J. P., Nimmo, E. R., Allshire, R. C. & Cech, T. R. Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature* **385**, 744–747 (1997).

75. Baumann, P. & Cech, T. R. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171–1175 (2001).

76. Loayza, D. & de Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* **423**, 1013–1018 (2003).

77. Colgin, L. M., Baran, K., Baumann, P., Cech, T. R. & Reddel, R. R. Human POT1 facilitates telomere elongation by telomerase. *Curr. Biol.* **13**, 942–946 (2003).

78. Hockemeyer, D., Daniels, J. P., Takai, H. & de Lange, T. Recent expansion of the telomeric complex in rodents: two distinct POT1 proteins protect mouse telomeres. *Cell* **126**, 63–77 (2006).

79. Wu, L. et al. Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell* **126**, 49–62 (2006).

80. Celli, G. B. & de Lange, T. DNA processing is not required for ATM-mediated telomere damage response after *TRF2* deletion. *Nature Cell Biol.* **7**, 712–718 (2005).

81. Wang, R. C., Smogorzewska, A. & de Lange, T. Homologous recombination generates T-loop-sized deletions at human telomeres. *Cell* **119**, 355–368 (2004).

82. Nugent, C. I. et al. Telomere maintenance is dependent on activities required for end repair of double-strand breaks. *Curr. Biol.* **8**, 657–660 (1998).

83. Tsukamoto, Y., Taggart, A. K. & Zakian, V. A. The role of the Mre11-Rad50-Xrs2 complex in telomerase-mediated lengthening of *Saccharomyces cerevisiae* telomeres. *Curr. Biol.* **11**, 1328–1335 (2001).

84. Takata, H., Tanaka, Y. & Matsuura, A. Late S phase-specific recruitment of Mre11 complex triggers hierarchical assembly of telomere replication proteins in *Saccharomyces cerevisiae*. *Mol. Cell* **17**, 573–583 (2005).

85. Nakamura, T. M., Moser, B. A. & Russell, P. Telomere binding of checkpoint sensor and DNA repair proteins contributes to maintenance of functional fission yeast telomeres. *Genetics* **161**, 1437–1452 (2002).

86. Takata, H., Kanoh, Y., Gunze, N., Shirahige, K. & Matsuura, A. Reciprocal association of the budding yeast ATM-related proteins Tel1 and Mec1 with telomeres *in vivo*. *Mol. Cell* **14**, 515–522 (2004).

87. Chan, S. W., Chang, J., Prescott, J. & Blackburn, E. H. Altering telomere structure allows telomerase to act in yeast lacking ATM kinases. *Curr. Biol.* **11**, 1240–1250 (2001).

88. Chan, S. W. & Blackburn, E. H. Telomerase and ATM/Tel1p protect telomeres from nonhomologous end joining. *Mol. Cell* **11**, 1379–1387 (2003).

89. Vaziri, H. et al. ATM-dependent telomere loss in aging human diploid fibroblasts and DNA damage lead to the post-translational activation of p53 protein involving poly(ADP-ribose) polymerase. *EMBO J.* **16**, 6018–6033 (1997).

90. Ranganathan, V. et al. Rescue of a telomere length defect of Nijmegen breakage syndrome cells requires NBS and telomerase catalytic subunit. *Curr. Biol.* **11**, 962–966 (2001).

91. Zhu, X. D., Kuster, B., Mann, M., Petruini, J. H. & de Lange, T. Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. *Nature Genet.* **25**, 347–352 (2000).

92. Verdun, R. E., Crabbe, L., Haggblom, C. & Karlseder, J. Functional human telomeres are recognized as DNA damage in G2 of the cell cycle. *Mol. Cell* **20**, 551–561 (2005).

93. Verdun, R. E. & Karlseder, J. The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres. *Cell* **127**, 709–720 (2006).

94. Cesare, A. J. & Griffith, J. D. Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops. *Mol. Cell. Biol.* **24**, 9948–9957 (2004).

95. Bailey, S. M., Brenner, M. A. & Goodwin, E. H. Frequent recombination in telomeric DNA may extend the proliferative life of telomerase-negative cells. *Nucleic Acids Res.* **32**, 3743–3751 (2004).

96. Bechter, O. E., Shay, J. W. & Wright, W. E. The frequency of homologous recombination in human ALT cells. *Cell Cycle* **3**, 547–549 (2004).

97. McEachern, M. J. & Haber, J. E. Break-induced replication and recombinational telomere elongation in yeast. *Annu. Rev. Biochem.* **75**, 111–135 (2006).

98. Lundblad, V. Telomere maintenance without telomerase. *Oncogene* **21**, 522–531 (2002).

99. Carson, C. T. et al. The Mre11 complex is required for ATM activation and the G2/M checkpoint. *EMBO J.* **22**, 6610–6620 (2003).

100. Haber, J. E. Partners and pathways repairing a double-strand break. *Trends Genet.* **16**, 259–264 (2000).

Acknowledgements We are indebted to V. Lundblad for constructive comments on the manuscript, the National Institutes of Health for funding (J.K.) and the Leukemia and Lymphoma Society for a long-term postdoctoral fellowship (R.E.V.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to J.K. (karseder@salk.edu).

Expandable DNA repeats and human disease

Sergei M. Mirkin¹

Nearly 30 hereditary disorders in humans result from an increase in the number of copies of simple repeats in genomic DNA. These DNA repeats seem to be predisposed to such expansion because they have unusual structural features, which disrupt the cellular replication, repair and recombination machineries. The presence of expanded DNA repeats alters gene expression in human cells, leading to disease. Surprisingly, many of these debilitating diseases are caused by repeat expansions in the non-coding regions of their resident genes. It is becoming clear that the peculiar structures of repeat-containing transcripts are at the heart of the pathogenesis of these diseases.

One of the central principles of classical (mendelian) genetics is that mutations are stably transmitted between generations. As long ago as 1918, however, a different type of inheritance was described for a human neurological disorder, myotonic dystrophy¹. This type of inheritance was characterized by increased expressivity: that is, a decreased age of onset and increased severity in individuals of subsequent generations. A similar hereditary pattern was later observed for other neurological diseases: for example, Huntington's disease, spinal and bulbar muscular atrophy, and several ataxias. The penetrance — that is, the probability that a given mutation results in disease — can also increase in successive generations, as was first demonstrated for fragile X syndrome². This unusual type of inheritance — characterized by a progressive increase in the expressivity and, sometimes, the penetrance of a mutation as it passes through generations — was called genetic anticipation.

Understanding this genetic anomaly became possible when the mutations that result in fragile X syndrome^{3,4} and spinal and bulbar muscular atrophy⁵ were characterized, which was soon followed by cloning of the gene that causes myotonic dystrophy^{6,7}. In all three cases, mutation seemed to arise from the continuous intergenerational expansion of simple trinucleotide repeats — (CNG)_n (where N denotes any nucleotide) — in different human genes. The progressive character of repeat expansion across generations provided a clue about the mechanism of genetic anticipation: the longer a repeat is, the more probable it is that it expands, and the more severe the phenotype. Thus, such mutations are classified as dynamic to account for the perpetual nature of the expansion process.

At present, expansions of simple DNA repeats are implicated in nearly 30 human hereditary disorders, and the list continues to grow (see refs 8 and 9 for recent updates). Various disease-causing repeats are depicted in the context of a fictitious human gene in Fig. 1. Most of these disorders are caused by the expansion of the triplet repeats (CGG)_n•(CCG)_n, (CAG)_n•(CTG)_n, (GAA)_n•(TTC)_n and (GCN)_n•(NGC)_n. But disease can also result from the expansion of the tetranucleotide repeat (CCTG)_n•(CAGG)_n, the pentanucleotide repeat (ATTCT)_n•(AGAAT)_n and even the dodecanucleotide repeat (C₄GC₄GCG)_n•(CGCG₄CG₄)_n. For a given hereditary disorder, only one repeat expands in a particular gene, strongly indicating that the molecular events leading to repeat expansions occur *in cis*. Expandable repeats can be located in various regions of their resident genes: first, the coding regions, as occurs in numerous diseases mediated by polyglutamine or polyalanine runs in proteins; second, the 5' untranslated regions (5'-UTRs), as in the case

of fragile X syndrome, fragile X mental retardation associated with the FRAXE site, fragile X tremor and ataxia syndrome, and spinocerebellar ataxia 12; third, 3'-UTRs, as is observed for myotonic dystrophy 1, spinocerebellar ataxia 8 and Huntington's-disease-like 2; fourth, introns, as in the case of myotonic dystrophy 2, Friedreich's ataxia and spinocerebellar ataxia 10; and fifth, promoter regions, as occurs in progressive myoclonic epilepsy 1.

Normal alleles of the genes associated with expansion-mediated diseases mostly contain either very short repetitive runs ('short-normal' alleles) or longer runs with several stabilizing interruptions ('long-normal' alleles): for example, AGG inserts within (CGG)_n runs in the gene associated with fragile X syndrome; CAT inserts within (CAG)_n runs in the gene associated with spinocerebellar ataxia 1; or GAG inserts within (GAA)_n runs in the gene associated with Friedreich's ataxia. Expansions begin when the length of an uninterrupted repetitive run exceeds a threshold of ~100–150 bases, often as a result of the loss of stabilizing interruptions at the end of the repetitive run in long-normal alleles^{10,11}. After this threshold is overcome, further expansions become progressively more likely, leading to the accumulation of dozens of repeats (for those that encode polyglutamine) to thousands of repeats (for those in non-coding regions) in just a few generations. Polyalanine-coding repeats behave differently from other expandable repeats¹². They are encoded by the imperfect triplet (GCN)_n. The threshold length for their expansion is extremely low (30–60 bases), and they rarely expand more than 1.5-fold. In addition, expanded (GCN)_n•(NGC)_n repeats are stable during both intergenerational and somatic transmission. Thus, progressive repeat lengthening, which is responsible for genetic anticipation, is not observed in polyalanine-mediated disorders.

This review concentrates on two questions. First, what are the mechanisms of repeat expansion? Studies carried out during the past decade suggest that expandable repeats are predisposed to instability, as a result of 'confusion' between the DNA replication, repair and recombination machineries (see refs 9, 13 and 14 for recent reviews). Second, how do repeat expansions result in disease? This question is particularly intriguing when considering the diseases that are caused by repeat expansions in the non-coding regions of human genes. It is becoming increasingly clear that a toxic 'gain of function' at the RNA level (see ref. 15 for a review) could be responsible. Here, I argue that the unusual structural characteristics of repetitive DNA and RNA, respectively, are central to these two issues.

¹Department of Biology, Tufts University, Medford, Massachusetts 02155, USA.

Molecular mechanisms of repeat expansions

Unusual structures of repetitive DNA

The first molecular model of how repeat expansions occur was based on DNA strand slippage during replication¹⁶. Looping out one or several repeats in the newly synthesized DNA strand should convert the loop into expansions after a second round of replication. However, this simple idea fails to explain why only a handful of all repeats expand and what determines the threshold length and the large-scale character of the expansions. The next breakthrough was the realization that all expandable repeats have unusual structural characteristics (see refs 13 and 17 for reviews). Single-stranded (CNG)_n repeats form hairpin-like structures that consist of both Watson–Crick base pairs and mismatched base pairs¹⁸ (Fig. 2a). In physiological conditions, the stability of such imperfect hairpins decreases according to the sequence of the triplet, in the order CGG>CTG>CAG=CCG, as a consequence of the energy contribution of the mismatched base pairs. Individual strands of (CCTG)_n•(CAGG)_n repeats have also been shown to fold into hairpin-like structures¹⁹. In addition to hairpins, single-stranded (CGG)_n, (CCG)_n and (CGCG₄CG₄)_n repeats can fold into tetrahedral structures^{20,21} stabilized by intertwining G quartets and i motifs (Fig. 2b).

The denaturation and renaturation of double-stranded DNA fragments that contain expandable repeats promote the formation of the ‘slipped-stranded’ DNA conformation. In this case, an out-of-register realignment of the complementary repetitive strands gives rise to ‘slip-outs’ that are folded into hairpin-like structures (Fig. 2c). These hairpins kinetically ‘trap’ repetitive DNA in the otherwise unfavourable slipped-stranded configuration²². Owing to the difference in hairpin-forming potential between expandable repeats in the complementary strands, slipped-stranded DNA is intrinsically asymmetrical. For example, when the (CTG)_n•(CAG)_n repeat converts into the slipped-stranded form, CAG slip-outs are mainly in the random-coil state, whereas CTG slip-outs are in the hairpin state²³. This asymmetry has important biological implications, because one of the complementary repetitive strands is usually more structure-prone than the other.

Slipped-stranded structures are not the only unusual structures formed by expandable repeats within double-stranded DNA. A (GAA)_n•(TTC)_n (homopurine–homopyrimidine) repeat can convert into an intramolecular triplex called H-DNA under the influence of negative supercoiling²⁴ (Fig. 2d). A different structure associated with the longer forms of this repeat is called sticky DNA²⁵. The main element of sticky DNA is a composite triplex (Fig. 2d), which is formed by the two distant, directly repeated (GAA)_n•(TTC)_n runs within circular

DNA²⁶; the exact configuration of the fourth repetitive strand remains to be elucidated. Lastly, an (A+T)-rich repeat that is responsible for spinocerebellar ataxia 10, (ATTCT)_n•(AGAAT)_n, belongs to the class of DNA-unwinding elements (Fig. 2e); that is, it unwinds progressively with increasing negative superhelical stress²⁷.

DNA replication models

It is generally thought that the unusual structural features of expandable repeats predispose them to instability. Indeed, repeats that are not structure-prone are considerably more stable genetically^{18,25,28}. Furthermore, the stabilizing effect of interruptions within the repetitive run in long-normal alleles is probably a result of their destabilizing effect on these unusual DNA structures^{18,29}. This has led to the idea that a misalignment between the two repetitive strands during DNA replication, further stabilized by unusual conformations of repetitive slip-outs (Fig. 3a), is the basis of repeat instability. After the next round of replication, either expansions or contractions occur, depending on the origin of the slipped-out strand³⁰. These slipped-stranded intermediates can be formed in the course of genetic processes that involve the separation of DNA strands: for example, DNA replication, repair and recombination. At present, each of these processes has been implicated in repeat expansions in one or another experimental system. It should be noted, however, that these unusual structures would be only transient intermediates during those processes, making their direct detection challenging.

Many models for repeat expansions assume that they occur during DNA replication for two main reasons. First, rapid accumulation of repetitive DNA cannot be explained without synthesis of massive amounts of DNA. Second, during the progression of the replication fork, a portion of the lagging-strand template that is known as the Okazaki initiation zone (OIZ) is always single stranded. The appearance of a repetitive run within this region could facilitate its folding into an unusual secondary structure.

Studies of DNA synthesis of expandable repeats *in vitro* support these ideas. Unusual DNA structures that are formed by expandable repeats during DNA synthesis *in vitro* stall various DNA polymerases^{20,24}. Occasionally, this stalling results in the misalignment of repetitive DNA strands, causing repeat expansions or contractions³¹.

In vivo, data from bacterial, yeast and mammalian cells show that the stability of expandable repeats depends substantially on their orientation relative to replication origins^{30,32–34}. Although these studies differ in important experimental details and interpretations, they generally agree

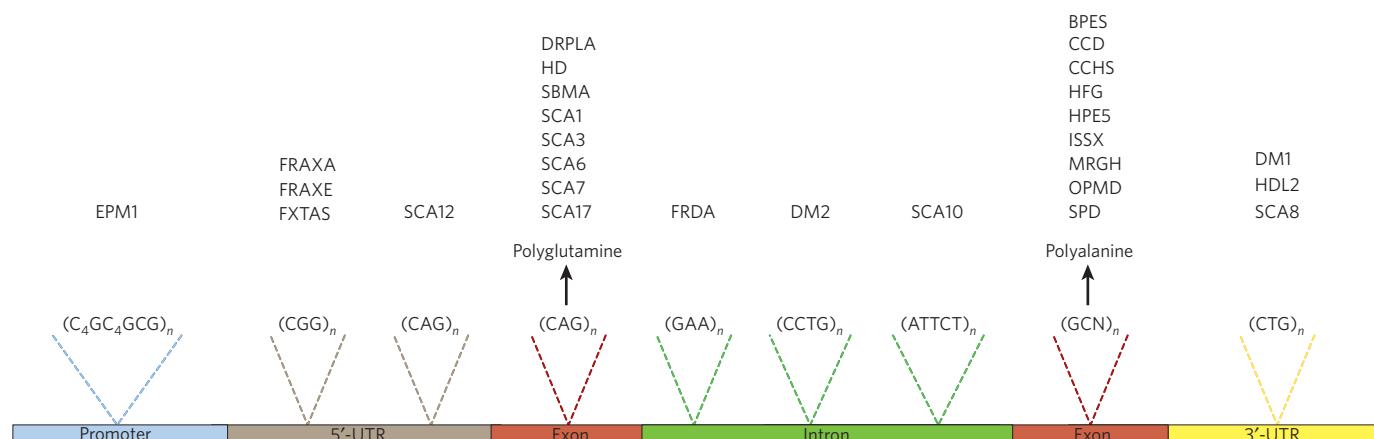


Figure 1 | Location of expandable repeats responsible for human diseases.

The sequence and location within a generic gene of expandable repeats that cause human diseases are shown, and the associated diseases are listed. BPES, blepharophimosis, ptosis and epicanthus inversus; CCD, cleidocranial dysplasia; CCHS, congenital central hypoventilation syndrome; DM, myotonic dystrophy; DRPLA, dentatorubral-pallidoluysian atrophy; EPM1, progressive myoclonic epilepsy 1; FRAXA, fragile X syndrome; FRAXE, fragile X mental retardation

associated with FRAXE site; FRDA, Friedreich’s ataxia; FXTAS, tremor and ataxia syndrome; HD, Huntington’s disease; HDL2, Huntington’s-disease-like 2; HFG, hand–foot–genital syndrome; HPE5, holoprosencephaly 5; ISSX, X-linked infantile spasm syndrome; MRGH, mental retardation with isolated growth hormone deficiency; OPMD, oculopharyngeal muscular dystrophy; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; SPD, synpolydactyly.

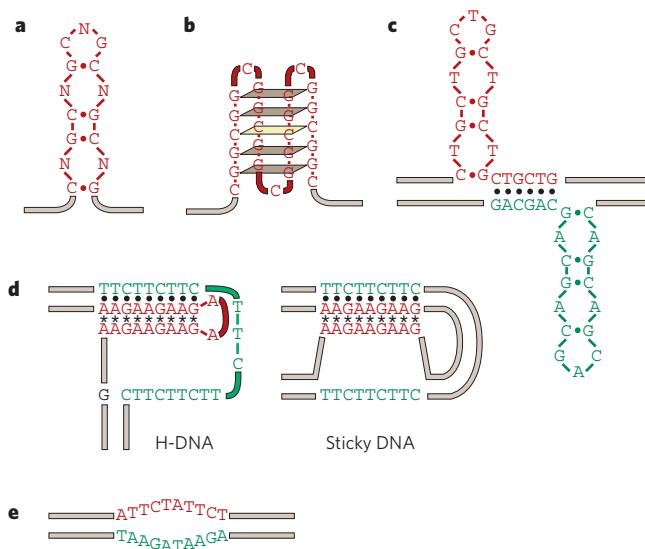


Figure 2 | Unusual DNA structures formed by expandable repeats.

Repetitive DNA can form several unusual structures, examples of which are shown. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige. **a**, An imperfect hairpin formed by $(\text{CNG})_n$ repeats. **b**, A quadruplex-like structure formed by the $(\text{CGG})_n$ repeat. Brown rectangles indicate G quartets, and the yellow rectangle indicates an i motif. **c**, A slipped-stranded structure formed by the $(\text{CTG})_n•(\text{CAG})_n$ repeat. **d**, H-DNA and sticky DNA formed by the $(\text{GAA})_n•(\text{TTC})_n$ repeat. Only one possible isoform, in which the homopurine strand is donated to the triplex, is shown for both structures. Reverse Hoogsteen pairing is indicated by asterisks. **e**, A DNA-unwinding element formed by the $(\text{ATTCT})_n•(\text{AGAAT})_n$ repeat.

that instability is the most marked when the structure-prone strand of a repetitive run functions as the lagging-strand template. The stability of repeats also depends on their distance from the replication origin^{34,35}. That is, the functioning of the structure-prone repetitive strand as the lagging-strand template and the precise location of the repetitive run in the OIZ are important factors in determining repeat instability.

Strong support for the role of DNA replication in repeat expansions came from studying yeast replication mutants. The frequencies of repeat expansions and contractions are affected markedly by mutations in several genes that encode proteins involved in replication: flap endonuclease (Fen1; also known as Rad27), DNA polymerase- δ , proliferating cell nuclear antigen, the large subunit of the clamp-loading complex, the helicase Srs2 (also known as Hpr5), and several other genes (see ref. 8 for a review). These proteins are involved in the synthesis of the lagging strand, the coordination between leading- and lagging-strand synthesis, and/or the restarting of the replication fork. Some of these mutations, most notably deletion of *FEN1*, destabilize various microsatellites and minisatellites. By contrast, others, such as deletion of *SRS2*, affect only expandable repeats³⁶. Recent genetic evidence from yeast studies suggests that the helicase Srs2 inhibits repeat expansions at a stage of post-replicative repair³⁷. It can therefore be concluded that some repetitive DNA intermediates, left behind after the replication fork passes on, can be converted into expansions if they remain unrepaired.

This view is in accord with direct observations of anomalous replication fork progression through expandable repeats in prokaryotic and eukaryotic cells^{38–40}. Various expandable repeats were found to stall the replication fork in all systems that have been studied. In almost all cases, replication stalling was evident when the length of the repetitive run approached the expansion threshold, and it was more marked when the structure-prone strand of the repetitive run was part of the lagging-strand template. In the region of the stall site, the lagging strand seemed to be under-replicated³⁸, implicating problems with the lagging-strand synthesis. Finally, repeats were particularly unstable in the orientations that were associated with replication stalling.

Together, these data led to the replication model for repeat instability (Fig. 3b), which is based on stalling and restarting of the replication fork (see ref. 14 for a review). In brief, formation of a stable secondary structure by a repetitive run in the lagging-strand template stalls lagging-strand synthesis and disrupts coordination with leading-strand synthesis. This could lead to synthesis continuing on the leading strand alone, while synthesis of the lagging strand resumes after skipping one or more Okazaki fragments, leaving a gap in the nascent lagging strand. Repeat contractions occur if a DNA polymerase involved in repair of the gap skips the structured portion of the lagging-strand template (Fig. 3b, upper pathway). Alternatively, replication stalling within a repetitive run can trigger replication fork reversal⁴¹. This would create a peculiar ‘chicken-foot’ structure with a single-stranded repetitive extension in the nascent leading strand, which can easily fold into a hairpin-like conformation. When the reversed replication fork is flipped back to restart replication, extra repeats can be added to the leading strand (Fig. 3b, lower pathway).

This model can account for several genetic features of repeat expansions at a molecular level. First, formation of unusual secondary structures in the lagging-strand template is more likely as the length of the repetitive run becomes comparable with the size of an OIZ. Thus, the similarity in expansion thresholds for various repeats might simply reflect the average size of the eukaryotic OIZ (~200 bases). Second, genetic anticipation could be explained in terms of consecutive replication stalls and restarts within longer repetitive runs, which would progressively increase their instability. Last, the disparity in the propensities of repeats to expand or contract in various model organisms could be explained by the differential probability of fork reversal and fork bypass in those organisms.

DNA-repair models

The unwinding of DNA during repair and recombination processes can also lead to the formation of slipped-stranded structures, which are implicated in repeat expansions (as mentioned earlier). The role of mismatch repair (MMR) in repeat instability has attracted particular attention. This is largely a result of studies using transgenic mouse models of Huntington’s disease and myotonic dystrophy. In these mice, mutational inactivation of the gene encoding MSH2 (MutS homologue 2) or MSH3 markedly decreased the frequency of repeat expansions during intergenerational transmission and in non-dividing somatic cells, shifting the pattern of repeat instability towards contractions^{42–45}. A heterodimer of MSH2 and MSH3 normally repairs single-base insertions and small loop-outs formed during DNA replication. So how could this complex promote, rather than prevent, repeat expansions? One possibility is that MSH2–MSH3 has an affinity for repetitive hairpins because these structures contain numerous mismatches. This interaction could therefore sequester MSH2–MSH3 to stabilize slipped-stranded intermediates instead of repairing them⁴⁵. In support of this idea, binding of MSH2–MSH3 to repetitive hairpins *in vitro* leads to discoordination of ATP hydrolysis and hairpin stabilization rather than repair⁴⁶.

How could this hijacking of MMR result in repeat expansions? Repeat expansions that occur during intergenerational transmission in humans and transgenic mice seem to happen in dividing cells about to undergo meiosis^{43,47,48}. As discussed earlier, replication fork stalling and restarting within a repetitive run might lead to the formation of a hairpin-like slip-out in the nascent DNA strand (Fig. 3b, bracketed intermediate). MSH2–MSH3 can be tricked into binding this hairpin through its similarity to mismatched DNA. Because the MMR machinery is thought to operate on nascent DNA strands, repetitive hairpins present in these strands would be stabilized preferentially over repetitive structures in template strands, shifting the equilibrium towards repeat expansions.

Repeat expansions are also observed in tissues in which cells do not divide, such as brain and skeletal muscle tissue in humans and mice^{49–52}. These events also require a functional MMR system^{42–45}, a somewhat unexpected finding given the lack of DNA replication in these cells. Furthermore, repeat expansions depend on the oxidative damage of

repetitive DNA in ageing non-dividing cells. Although expansions can occur during the repair of DNA nicks or gaps generated directly by oxygen radicals⁴⁵, recent studies using a transgenic mouse model of Huntington's disease have implicated base-excision repair as central to this process⁵³. Strikingly, in these mice, age-dependent repeat expansions in somatic cells depended on a single base-excision repair enzyme, 8-oxoguanine DNA glycosylase (OGG1) (see page 941). Removal of an oxidized guanine by OGG1 generates a nick in the repetitive run (Fig. 4a), and DNA-repair synthesis is then needed to heal this lesion. During this repair synthesis, the non-template DNA strand is displaced, forming a flap (Fig. 4b). Normally, a flap is removed by FEN1; this endonuclease is loaded onto the 5' end of the flap, migrates to the junction with the duplex, and cleaves the flap. If a flap contains the structure-prone strand of a repetitive run, however, it can fold into a hairpin-like conformation, complicating FEN1 loading^{54,55}. MSH2–MSH3 can further stabilize this hairpin (as discussed earlier), preventing flap removal (Fig. 4c). Completion of the repair process will yield a stable slipped-stranded DNA intermediate with a repeat extension in the nicked strand (Fig. 4d).

Can these unprocessed flaps be converted into expansions in non-dividing cells? This was indeed the case in a study of the fate of various slipped-stranded DNA intermediates in extracts of terminally differentiated neuron-like cells⁵⁶. Intermediates with repetitive slip-outs in the

nicked DNA strand (Fig. 4d) were repaired to become a set of products with differentially expanded repeats (Fig. 4e), as though the hairpin was incompletely excised during the repair synthesis. This 'error-prone' repair might therefore account for the final stages of repeat expansions in non-dividing cells.

Unprocessed flap structures can also contribute to repeat expansions in dividing cells, providing an elegant explanation of the bias towards repeat expansions that has been observed in human pedigrees. In support of this, inefficient flap removal leads to repeat expansions in yeast⁵⁷. The situation is less clear in mice, however, because (CTG)_n repeat stability was essentially unaltered in *Fen1*-knockout mice⁵⁸.

DNA recombination models

Various pathways of genetic recombination can also contribute to repeat instability. The simplest mechanism could be unequal crossing-over between the repetitive runs on homologous chromosomes during meiosis, resulting in reciprocal expansions and contractions (Fig. 5a). This process has been implicated in expansions of (GCN)_n repeats, which encode polyalanine⁵⁹. For polyalanine-mediated disorders, every allele in a given pedigree has sequence variations due to the redundancy in the third position of the repeat. Tracking these variations in alleles containing repeat expansions leads to the conclusion

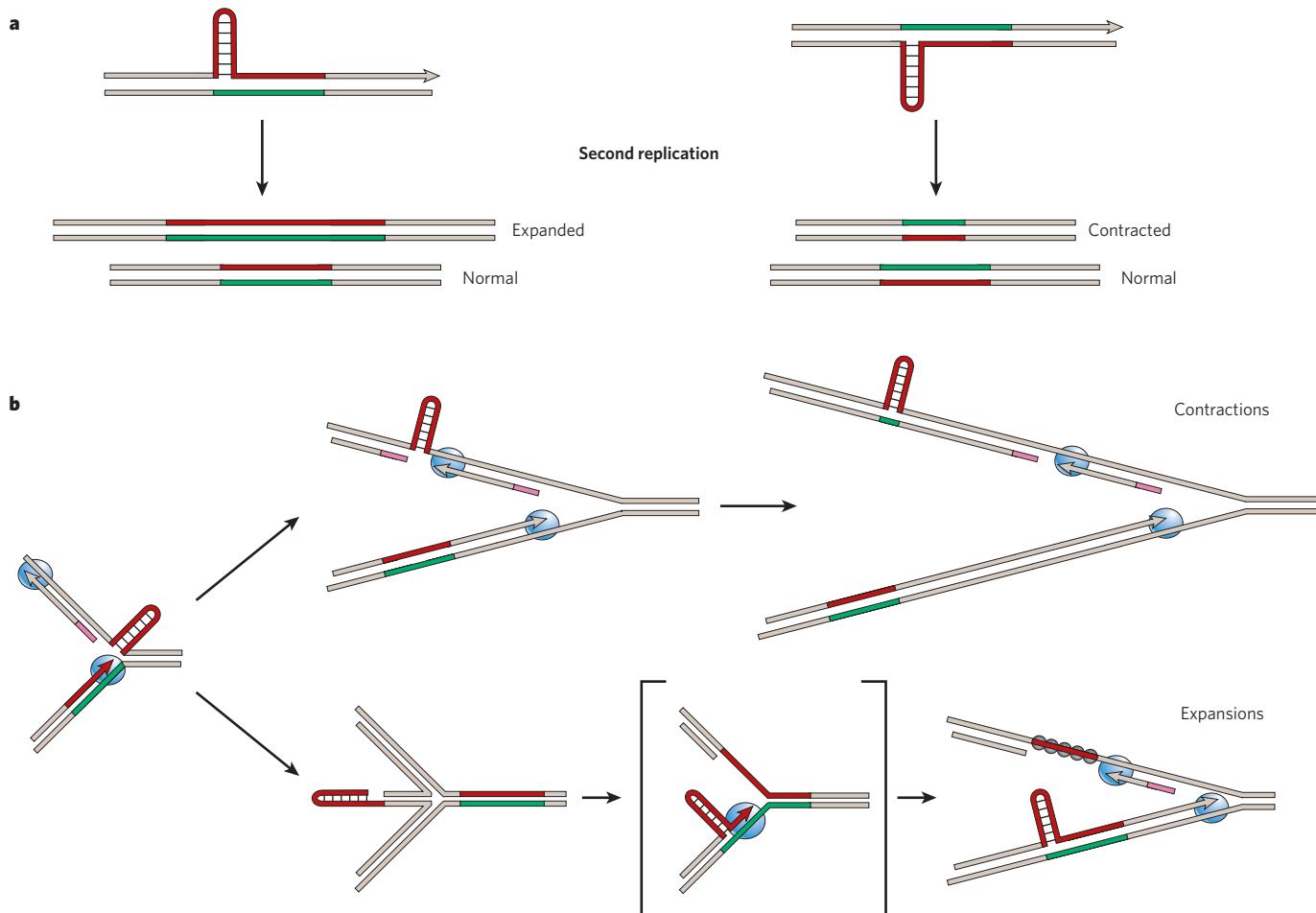


Figure 3 | Replication mechanisms for repeat expansion. **a**, After two rounds of replication, formation of a repetitive hairpin on the nascent strand results in repeat expansions (left panel), whereas the presence of the same structure on the template strand results in repeat contractions (right panel). **b**, A model for repeat instability based on replication fork stalling and restarting within the repetitive run is shown. Repeat contractions (upper pathway) occur when the machinery for the lagging-strand synthesis skips the repetitive hairpin on the lagging-strand template. Repeat expansions

(lower pathway) can occur during replication fork reversal and restart, leading to the formation of a repetitive hairpin on the nascent leading strand. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige. DNA polymerases are shown in blue, primers for Okazaki fragments in pink, and single-stranded-DNA-binding proteins as grey circles. The bracketed intermediate contains a hairpin on the nascent strand, which can also be stabilized by MSH2–MSH3.

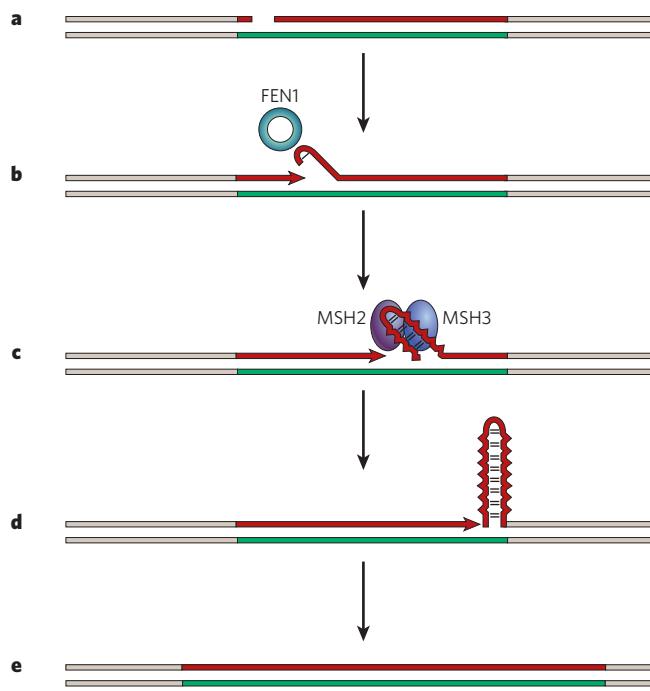


Figure 4 | Gap repair model for repeat expansions in non-dividing cells.

a, Oxidizing radicals generate a small gap in the structure-prone strand of a repetitive run. **b**, The loading of FEN1 onto a repetitive flap generated during the DNA-repair synthesis is impaired by hairpin formation. **c**, The binding of MSH2–MSH3 stabilizes the repetitive hairpin, preventing flap removal. **d**, A stable slipped-stranded DNA intermediate is formed on the completion of the repair synthesis. **e**, The slipped-stranded intermediate is converted into an expansion by an error-prone repair pathway. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige.

that the repeat expansions resulted from unequal crossing-over between (GCN)_n blocks in normal alleles.

Meiotic crossing-over has, however, been ruled out as the source of instability for all other repeats, because these expansions are not accompanied by an exchange of the flanking markers⁶⁰. This leaves mitotic recombination as the most plausible source. Indeed, expandable repetitive runs stimulate recombination in mitotically dividing cells, undergoing length changes during this process. In bacteria, (CAG)_n (ref. 61), (GAA)_n (ref. 62) and (CCTG)_n (ref. 63) repeats increase the rate of both inter- and intramolecular plasmid recombination. This increase is proportional to the length of the repetitive run, being particularly marked when a structure-prone repetitive strand functions as the lagging-strand template, pointing to the role of DNA replication (discussed later). Furthermore, expanded and contracted versions of these repeats were frequently detected among the recombination products. In yeast, long (CAG)_n•(CTG)_n runs were shown to cause chromosomal breakage, triggering ectopic recombination⁶⁴. They also stimulated spontaneous unequal sister-chromatid exchange⁶⁵ and underwent frequent expansions and contractions during gene conversion^{66,67}. Finally, in mitotically dividing mammalian cells, (CAG)_n•(CTG)_n repeats were shown to stimulate homologous recombination, undergoing contractions and other rearrangements during this process⁶⁸.

To account for the recombinogenic activity of expandable repeats, it is safe to assume that, in one way or another, the repeats trigger the formation of double-strand breaks in DNA. Notably, data from bacteria and yeast show that stimulation of recombination depends on the orientation of repeats within a replicon, pointing to the connection between their replication and recombination. One possible mechanism of such a connection is presented in Fig. 5b. As discussed earlier, the formation of a stable secondary structure in the lagging-strand template by an

expandable repeat leads to stalling of the replication fork. Cleavage and processing of this structure by an unidentified eukaryotic endonuclease — a functional homologue of the bacterial nuclease SbcCD — would generate a DNA fragment with a single-stranded 3' repetitive extension capable of invading the sister chromatid. An out-of-register invasion would then lead to repeat expansions or contractions (Fig. 5b). Another plausible recombinogenic intermediate could be the product of fork reversal (Fig. 3b), which contains a single-stranded repetitive run at the 3' end of the leading strand.

It is worth noting that single-stranded repetitive tails (discussed earlier) can also invade homologous repeats at ectopic positions. Subsequent recombination events would lead to chromosomal rearrangements, such as translocations and deletions. Expandable repeats have indeed been shown to induce gross chromosomal rearrangements (see ref. 69 for a review).

Initial stages

All of the molecular models described here considered repeats beyond the expansion threshold. But what happens at the initial stages, when a normal allele converts into the pre-mutation allele capable of expanding in future generations? These early events could involve either the accidental lengthening of short-repetitive runs in normal alleles or the loss of stabilizing interruptions in long-normal alleles. Most of the data point to the second possibility.

Two prospective mechanisms leading to the loss of stabilizing interruptions are presented in Fig. 6. An interruption that has slipped out of the template strand during DNA replication would simply be lost if this slippage were unrepaired (Fig. 6a). This simple model cannot, however, account for two observations: first, more than one interruption is usually lost en route from the long-normal allele to its expandable counterpart; and, second, expansions in genetic carriers usually occur at one end of a repetitive run^{10,11}. An elegant explanation that has emerged from yeast studies is presented in Fig. 6b. Expandable repeats in yeast are stabilized by interruptions, as is the case in humans. This stable maintenance seems to require the MMR system, because knockdown of expression of the genes involved in MMR led to the frequent loss of interruptions. This stabilization is explained by the ‘co-excision’ mechanism⁷⁰. Formation of a slipped-stranded intermediate during replication of a repetitive run that contains several interruptions would place these interruptions out of register in both the hairpin and the duplex part of this intermediate. Co-excision of the hairpin and the mismatches in a duplex part of such an intermediate is needed to maintain the original sequence of the repeat. Failure to do so would generate a non-interrupted expansion at one end of the repeat after another round of replication.

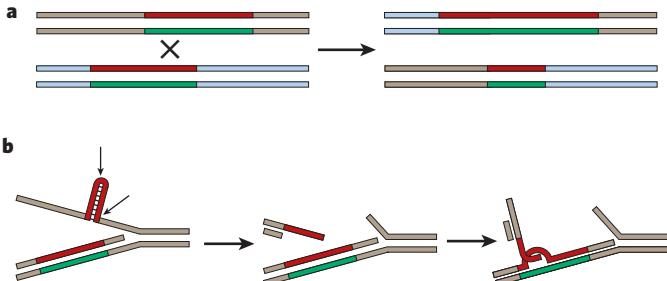


Figure 5 | Recombination models for repeat expansions. **a**, Unequal crossing-over results in the reciprocal appearance of expanded and contracted repeats. Homologous chromosomes are shown in beige and blue, and repetitive DNA strands are shown in red and green. **b**, Cleavage of a stable DNA structure on the lagging-strand template (left), formed during DNA replication, generates a single-stranded 3' repetitive extension (centre). Out-of-register invasion of such an extension into a sister chromatid (right) might lead to repeat expansions or contractions. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige. Small arrows show potential cleavage sites.

As discussed earlier, formation of slipped-stranded replication intermediates depends on the mode of replication fork progression through the repeat: that is, its orientation relative to the replication origin, as well as its exact position in the lagging-strand template. These considerations led to three hypotheses, linking early expansion stages with the position and orientation of the repetitive run in the replication unit. First, a hypothesis known as 'ori-switch'⁷¹ suggests that inactivation of the replication origin on one side of a repeat, combined with the activation of a cryptic origin on its other side, triggers expansions by placing the structure-prone strand of the repetitive run as the lagging-strand template (Fig. 7a). Second, a hypothesis known as 'ori-shift'⁷¹ assumes that early stages of expansions depend on the position of a repeat within the OIZ; therefore, a change in the distance between the replication origin and the repeat (caused by insertion of a mobile element, for example) could induce expansions (Fig. 7b). Third, a hypothesis known as the fork-shift model⁷² proposes that a change in the mode of replication fork progression, caused by an epigenetic event in the vicinity of the repeat, could alter the position of this repeat within the OIZ, leading to expansions (Fig. 7c). These models remain to be substantiated by fine analysis of replication fork progression through expandable repeats in cells from normal and affected individuals. However, a particularly intriguing possibility is that the predicted changes could depend on the developmental or tissue-specific mode of origin usage in mammals.

Molecular mechanisms of repeat-mediated RNA toxicity

Most diseases associated with repeat expansions show dominant inheritance. Classical genetics explains dominance through the effects of mutations on protein function. Those effects include loss of function, leading to haploinsufficiency, and gain of function, when a mutant protein has a dominant-negative effect on its normal counterpart or acquires a novel, deleterious function. This explanation seems to account for diseases caused by repeat expansions in coding sequences. In the case of polyglutamine- or polyalanine-mediated disorders, mutant proteins acquire a deleterious ability to aggregate, which might trigger cell death and tissue degeneration (see ref. 73 for a review).

Unexpectedly, however, the expression of expanded repeats in non-coding sequences also gives rise to dominant mutations (see ref. 15 for a review). This is true for myotonic dystrophy 1 and 2, spinocerebellar ataxia 8, 10 and 12, Huntington's-disease-like 2, and fragile X tremor and ataxia syndrome. These diseases are caused by various tri-, tetra- and pentanucleotide repeats, which are situated in the various non-coding regions (5'-UTRs, 3'-UTRs and introns) of their resident genes (Fig. 1). The scale of expansions also differs between these diseases, from as many as 11,000 repeats for myotonic dystrophy 2 to fewer than 100 repeats for Huntington's-disease-like 2 and spinocerebellar ataxia 8.

Recent progress in the field indicates that these dominantly inherited diseases could be caused by gain of function at the RNA level. Originally, this idea came from studies of myotonic dystrophy. First, transcripts of *DMPK* (which encodes myotonic dystrophy protein kinase 1) containing expanded CUG repeats were shown to be retained in the nuclei of fibroblasts and myoblasts, forming distinct foci⁷⁴. Second, the ability of normal myoblasts to undergo myogenic differentiation in cell culture seemed to be suppressed by overexpression of RNA containing a (CUG)₂₀₀ repeat⁷⁵. Most strikingly, a transgenic mouse expressing (CUG)₂₅₀ repeats within the 3'-UTR of a heterologous gene (skeletal actin) showed major symptoms of myotonic dystrophy⁷⁶.

Similar, somewhat less compelling, observations have begun to accumulate for other diseases. In patients with fragile X tremor and ataxia syndrome, transcripts containing intermediate-size (CGG)_n repeats are overproduced, forming intranuclear inclusions in the neurons and astrocytes⁷⁷. Furthermore, overexpression of (CGG)_n-containing transcripts leads to the appearance of intranuclear inclusions, together with neurodegeneration, in mice and *Drosophila melanogaster*⁷⁸. A role for 'toxic' RNA is also favoured as a possibility in the development of spinocerebellar ataxia 8, because transcription of the human SCA8 gene containing an expanded (CTG)_n repeat in the retina of *D. melanogaster* causes its neurodegeneration⁷⁹. In addition, overexpression of the repeat associated with spinocerebellar ataxia 10 leads to the formation of intranuclear foci containing (AUUCU)_n transcripts in cell culture⁸⁰. Finally, there are promising data that point to the existence of similar mechanisms for the development of spinocerebellar ataxia 12 and Huntington's-disease-like 2.

RNA gain-of-function effects could be grounded in the unusual structural features of repeat-containing RNAs (Fig. 8A). The earliest support for this idea came from the observation that (CUG)_n repeats in the natural sequence context of the *DMPK* transcript formed imperfect, mismatched hairpins, the stability of which increased with the length of the repeat⁸¹. This study was subsequently extended to all 20 possible triplet repeats in RNA⁸². It seems that six repetitive motifs — (CGU)_n, (CGA)_n, (CAG)_n, (CUG)_n, (CCG)_n and (CGG)_n — can form stable RNA hairpins. The stability of RNA hairpins that consist of (CNG)_n, which are implicated in diseases associated with repeat expansions, was found to depend on the nature of mismatched base pairs, decreasing in the order of CGG>CUG>CCG>CAG. A tetranucleotide repeat involved in myotonic dystrophy 2, (CCUG)_n, was also shown to form an RNA hairpin with twice as many mismatches. Finally, a pentanucleotide repeat responsible for spinocerebellar ataxia 10, (AUUCU)_n, folded into an unusual RNA hairpin-like structure stabilized by non-Watson–Crick A•U and U•U base pairs⁸³. Interestingly, stabilizing interruptions, such as AGG inserts within long-normal (CGG)_n-containing alleles, led to the formation of branched RNA hairpin structures⁸⁴ (Fig. 8A, c). It

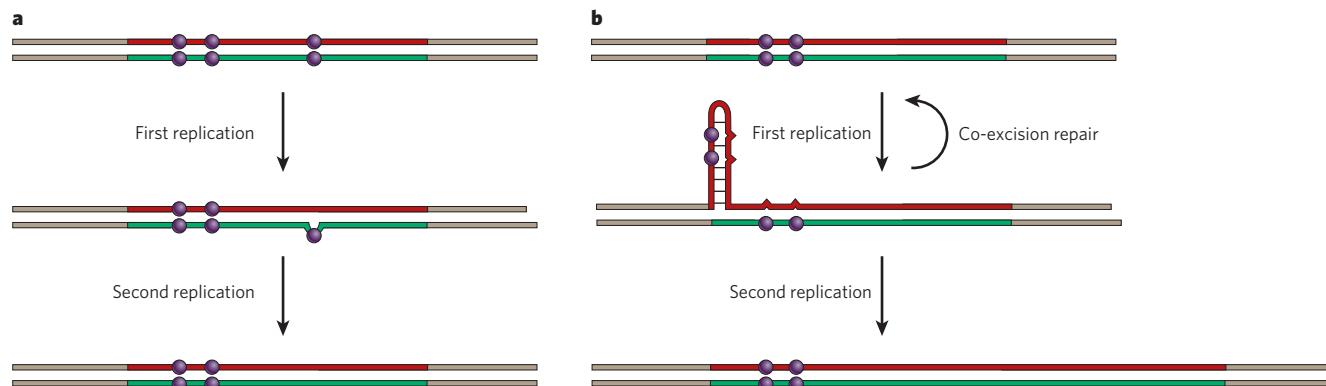


Figure 6 | Loss of stabilizing interruptions within expandable repeats.

Two possible mechanisms for the loss of stabilizing interruptions. **a**, An interruption (purple circle) that has slipped out of the template strand (green) is lost during DNA replication. **b**, Misalignment of nascent and template DNA strands in long-normal alleles creates mismatches in both

the hairpin and duplex part of the slipped-stranded structure during replication. These mismatches can be repaired by co-excision repair. Failure to repair them leads to expansions at the 3' end of the repetitive run. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige.

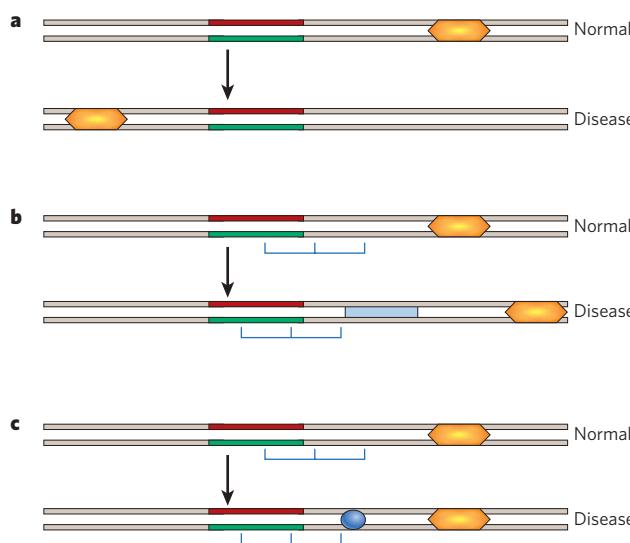


Figure 7 | Three models of early events in repeat expansions. Three hypotheses have been put forward to explain early events in repeat expansions: the ori-switch model, the ori-shift model and the fork-shift model. **a**, The ori-switch model proposes that the initial event leading to repeat expansion is a reversal in the direction of replication (replication origin shown in orange) through the repetitive run, so the structure-prone strand of repetitive run becomes the lagging-strand template. **b**, The ori-shift model assumes that a change in the distance between the replication origin and the repetitive run (for example, as a result of insertion of a mobile element; blue rectangle) changes the position of the repetitive run within the OIZ (consecutive OIZs shown as blue brackets), eliciting repeat expansions. **c**, The fork-shift model proposes that an expansion-prone change in the position of the repetitive run within the OIZ is triggered by an epigenetic event (blue oval) that does not affect the orientation of a repetitive run or its distance from the replication origin. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige.

was therefore suggested that this branching precludes repetitive RNA encoded by long-normal alleles from becoming toxic.

How these RNA structures can cause gain of function and disease remains to be determined. The most intensively discussed potential mechanisms can be referred to as protein sequestration, RNA degradation and chromatin silencing (Fig. 8B). Although there is considerable experimental evidence for the protein sequestration model, the other two mechanisms are only beginning to be tested.

The main support for the protein sequestration model comes from studies of myotonic dystrophy 1 and 2. Intranuclear foci that are characteristic of these diseases contain at least seven RNA-binding proteins associated with (CUG)_n- and (CCUG)_n-containing transcripts, including three Muscleblind-like (MBNL)-family proteins^{85,86} and two different CUG RNA-binding proteins (CUG-BPs)⁸⁷. It was suggested that sequestration of these proteins could lead to RNA gain of function if these proteins are required for the normal expression of muscle-, heart- and brain-specific genes. It is becoming increasingly clear that a key molecular event leading to myotonic dystrophy is the deregulation of alternative RNA splicing during development⁸⁸. At least 13 splicing events are disturbed in muscle, heart and brain tissues from patients with myotonic dystrophy, and an embryonic ‘blueprint’ for splicing is almost always retained at the expense of the adult splicing pattern (see ref. 89 for a review). Both MBNL1 and CUG-BP1 are implicated in these splicing events, and the splicing pattern characteristic for myotonic dystrophy is consistent with the loss of MBNL1 and gain of CUG-BP1 activities^{88,90}.

Although the fine details of the interplay between MBNL1 and CUG-BP1 remain to be understood, most of the data suggest that MBNL1 sequestration by long mismatched hairpins (Fig. 8B, a) is a key event leading to splicing deregulation and, eventually, disease. MBNL1

binds strongly to (CUG)_n and (CCUG)_n hairpins, but not to perfect (CUG)_n•(CAG)_n RNA duplexes, implicating a role for non-Watson–Crick U•U mismatches⁹¹. Most importantly, *Mbnl1*-knockout mice show splicing deregulation and phenotypic manifestations characteristic of myotonic dystrophy, such as skeletal muscle myotonia and cataracts⁹². Finally, the myotonic-dystrophy-like phenotype of mice expressing (CUG)_n from a heterologous gene⁷⁶ could be reversed by overproduction of MBNL1 (ref. 93).

Protein sequestration has also been implicated in other diseases caused by repeat expansions in non-coding RNA. For example, MBNL1 is present in RNA foci that have been observed in cells isolated from patients with fragile X tremor and ataxia syndrome⁹⁴. Furthermore, neurodegeneration caused by expression of a mutant human SCA8 gene in *D. melanogaster* retina is modified by mutations in the *D. melanogaster* muscleblind gene⁷⁹. This model, however, seems more plausible for the diseases that are caused by massive repeat expansions, where exceptionally long RNA hairpins could efficiently sequester these RNA-binding proteins, rather than for diseases with moderate-sized repeat expansions.

What could be an alternative mechanism for RNA gain of function? One exciting possibility is that repetitive RNA hairpins are reminiscent of unprocessed microRNAs. Processing of such hairpins by the RNA interference (RNAi) pathway could lead to silencing of genes that contain short complementary repeats in their transcripts, resulting in disease⁹⁵. This hypothesis, the RNA degradation model (Fig. 8B, b), was initially supported by observations that (CGG)_n hairpins are ‘digested’, albeit inefficiently, by the human protein Dicer, an RNase that is central to the

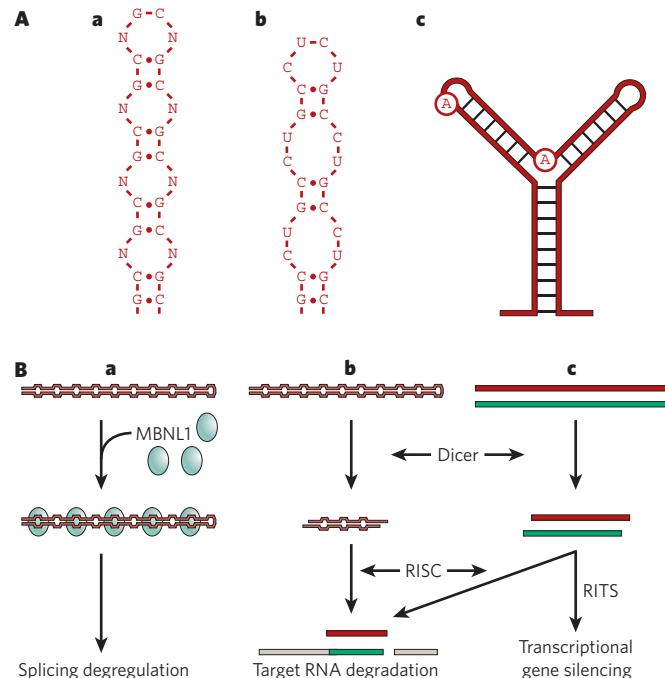


Figure 8 | Disease-associated RNA gain of function. **A**, Mismatched RNA hairpins can be formed by various expandable repeats: (CNG)_n (**a**), (CCUG)_n (**b**), and (CGG)_n with two stabilizing AGG interruptions, with the A encircled (**c**). **B**, Molecular models of RNA gain of function. **a**, A repetitive RNA hairpin sequesters the protein MBNL1, which is required for normal splicing. **b**, A repetitive RNA hairpin is cleaved by Dicer and processed by the RNA-induced silencing complex (RISC), leading to the degradation of transcripts carrying short complementary repeats. **c**, Bidirectional transcription through repetitive runs generates double-stranded RNA. This leads to either target RNA degradation, if the processing is carried out by RISC, or chromatin silencing, if the processing is carried out by the RNA-induced transcriptional silencing complex (RITS). The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige.

RNAi response⁹⁶. Recently, a variety of (CNG)_n-containing transcripts were shown to be targets of Dicer both *in vitro* and *in vivo*. Remarkably, the resultant short (CNG)_n repeats functioned as siRNAs and triggered downstream silencing effects⁹⁷.

Another intriguing model involves chromatin silencing (Fig. 8B, c). This model is derived from recent data showing that in some diseases (such as myotonic dystrophy 1 and spinocerebellar ataxia 8), transcription across expandable repeats can proceed in both directions^{98,99}. In the case of myotonic dystrophy 1, transcription of both the sense and antisense strands of the repetitive run led to the formation of a 21-nucleotide duplex RNA, pointing to the involvement of RNAi mechanisms. This antisense transcription was also linked to methylation of the lysine residue at position 9 of histone H3 and to recruitment of heterochromatin protein 1 (that is, to local formation of heterochromatin). Such RNA-induced initiation of transcriptional gene silencing could affect neighbouring gene expression, accounting for the diverse clinical manifestations of myotonic dystrophy 1 and perhaps of other diseases. More studies are needed to establish whether, and to what extent, different RNAi pathways are involved in repeat-expansion-associated diseases.

Future directions

Although the mechanisms that are responsible for repeat expansions in DNA are generally understood, many important questions remain unanswered. For example, will we find more expandable repeats and associated diseases in the future? If the structural concept presented in this review is correct, one would expect other structure-forming repeats to expand. But which factors are responsible for the initial expansions in human pedigrees? Mapping the replication origins in the vicinity of expandable repeats and studying in detail the mode of replication fork progression through the repeats in various cells and tissues should shed light on this matter. For example, recent mapping of the replication origin at the fragile X locus (*FRAXA*) provides additional support for the replication model of repeat expansions by showing that the structure-prone strand of the repetitive run functions as the lagging-strand template¹⁰⁰.

Another question is whether mechanisms of repeat expansion are conserved among various organisms or even among different cell types in the same organism. There are marked differences in expansion and contraction biases between mammals and unicellular organisms: repeats readily expand in humans but mostly contract in bacteria and yeast. In addition, repeats differ in their propensity to expand between various human cell types. We need to understand the reasons for these differences if we are to develop therapeutic approaches that could induce repeat contractions in humans. Could the trend for expansion of repeats in humans be reversed by pharmaceutical intervention? On the basis of our current knowledge, studies of drugs that affect DNA replication, repair or cell-cycle checkpoints are warranted.

Even more questions remain when it comes to the molecular pathways that lead from repeat expansions to disease. How widespread is the involvement of RNA toxicity in disease development? What is the role of repeat-containing RNA in the pathogenesis of polyglutamine-mediated disorders? What is the role of RNAi, broadly defined, in repeat-expansion-associated diseases? RNAi might have a considerable role, given that it provides an elegant explanation for the observed genetic dominance. And does chromatin silencing have a role in repeat expansion diseases? This emerging research direction seems to hold promise. Finally, is it possible to inhibit the aberrant processing of repeat-containing RNAs? If so, this would be of prime medical importance. Studies aimed at addressing these issues and related questions are underway in many laboratories worldwide.

1. Fleischer, B. Über myotonische Dystrophie mit Katarakt. *Albrecht Von Graefes Arch. Klin. Exp. Ophthalmol.* **96**, 91–133 (1918).
2. Sherman, S. L. *et al.* Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum. Genet.* **69**, 289–299 (1985).
3. Verkerk, A. J. *et al.* Identification of a gene (*FMR-1*) containing a CCG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914 (1991).
4. Kremer, E. J. *et al.* Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG). *Science* **252**, 1711–1714 (1991).
5. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**, 77–79 (1991).
6. Brook, J. D. *et al.* Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**, 799–808 (1992).
7. Mahadevan, M. *et al.* Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**, 1253–1255 (1992).
8. Mirkin, S. M. Molecular models for repeat expansions. *Chemtracts Biochem. Mol. Biol.* **17**, 639–662 (2004).
9. Pearson, C. E., Nichol Edamura, K. & Cleary, J. D. Repeat instability: mechanisms of dynamic mutations. *Nature Rev. Genet.* **6**, 729–742 (2005).
10. Kunst, C. B. & Warren, S. T. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* **77**, 853–861 (1994).
11. Jodice, C. *et al.* Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia I. *Am. J. Hum. Genet.* **54**, 959–965 (1994).
12. Brown, L. Y. & Brown, S. A. Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet.* **20**, 51–58 (2004).
13. Wells, R. D., Dere, R., Hebert, M. L., Napierala, M. & Son, L.S. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* **33**, 3785–3798 (2005).
14. Mirkin, S. M. DNA structures, repeat expansions and human hereditary disorders. *Curr. Opin. Struct. Biol.* **16**, 351–358 (2006).
15. Ranum, L. P. & Cooper, T. A. RNA-mediated neuromuscular disorders. *Annu. Rev. Neurosci.* **29**, 259–277 (2006).
16. Kunkel, T. A. Slippery DNA and diseases. *Nature* **365**, 207–208 (1993).
17. McMurray, C. T. DNA secondary structure: a common and causative factor for expansion in human disease. *Proc. Natl. Acad. Sci. USA* **96**, 1823–1825 (1999).
18. Gacy, A. M., Goellner, G., Juranic, N., Macura, S. & McMurray, C. T. Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell* **81**, 533–540 (1995).
19. Dere, R., Napierala, M., Ranum, L. P. & Wells, R. D. Hairpin structure-forming propensity of the (CCTG•CAGG) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2. *J. Biol. Chem.* **279**, 41715–41726 (2004).
20. Usdin, K. & Woodford, K. J. CCG repeats associated with DNA instability and chromosome fragility from structures that block DNA synthesis *in vitro*. *Nucleic Acids Res.* **23**, 4202–4209 (1995).
21. Fry, M. & Loeb, L. A. The fragile X syndrome d(CGG)_n nucleotide repeats form a stable tetrahedral structure. *Proc. Natl. Acad. Sci. USA* **91**, 4950–4954 (1994).
22. Pearson, C. E. & Sinden, R. R. Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry* **35**, 5041–5053 (1996).
23. Pearson, C. E. *et al.* Slipped-strand DNAs formed by long (CAG)•(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucleic Acids Res.* **30**, 4534–4547 (2002).
24. Gacy, A. M. *et al.* GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol. Cell* **1**, 583–593 (1998).
25. Sakamoto, N. *et al.* Sticky DNA: self association properties of long (GAA)•(TTC) repeats in R•R•Y triplex structures from Friedreich's ataxia. *Mol. Cell* **3**, 465–475 (1999).
26. Vetcher, A. A. *et al.* Sticky DNA, a long (GAA•GAA•TTC) triplex that is formed intramolecularly, in the sequence of intron 1 of the frataxin gene. *J. Biol. Chem.* **277**, 39217–39227 (2002).
27. Potaman, V. N. *et al.* Unpaired structures in SCA10 (ATTCT)_n•(AGAAT)_n repeats. *J. Mol. Biol.* **326**, 1095–1111 (2003).
28. Moore, H., Greenwell, P. W., Liu, C. P., Arnheim, N. & Petes, T. D. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc. Natl. Acad. Sci. USA* **96**, 1504–1509 (1999).
29. Sakamoto, N. *et al.* GGA•TCC-interrupted triplets in long GAA•TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. *J. Biol. Chem.* **276**, 27178–27187 (2001).
30. Kang, S., Jaworski, A., Ohshima, K. & Wells, R. D. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nature Genet.* **10**, 213–218 (1995).
31. Ohshima, K. & Wells, R. D. Hairpin formation during DNA synthesis primer realignment *in vitro* in triplet repeat sequences from human hereditary disease genes. *J. Biol. Chem.* **272**, 16798–16806 (1997).
32. Freudenreich, C. H., Stavenhagen, J. B. & Zakian, V. A. Stability of a CTG•CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol.* **17**, 2090–2098 (1997).
33. Miret, J. J., Pessoa-Brandao, L. & Lahue, R. S. Orientation-dependent and sequence-specific expansions of CTG•CAG trinucleotide repeats in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **95**, 12438–12443 (1998).
34. Cleary, J. D., Nichol, K., Wang, Y. H. & Pearson, C. E. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nature Genet.* **31**, 37–46 (2002).
35. Rindler, M. P., Clark, R. M., Pollard, L. M., De Biase, I. & Bidichandani, S. I. Replication in mammalian cells recapitulates the locus-specific differences in somatic instability of genomic GAA triplet-repeats. *Nucleic Acids Res.* **34**, 6352–6361 (2006).
36. Bhattacharyya, S. & Lahue, R. S. *Saccharomyces cerevisiae* Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. *Mol. Cell. Biol.* **24**, 7324–7330 (2004).
37. Daee, D. L., Mertz, T., Collins, N. & Lahue, R. S. Post-replication repair inhibits CAG•CTG repeat expansions in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**, 102–110 (2007).
38. Samadashwily, G. M., Raca, G. & Mirkin, S. M. Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.* **17**, 298–304 (1997).
39. Krasilnikova, M. M. & Mirkin, S. M. Replication stalling at Friedreich's ataxia (GAA)_n repeats *in vivo*. *Mol. Cell. Biol.* **24**, 2286–2295 (2004).

40. Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R. S. & Mirkin, S. M. Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol.* **23**, 1349–1357 (2003).

41. Fouché, N., Ozgur, S., Roy, D. & Griffith, J. D. Replication fork regression in repetitive DNAs. *Nucleic Acids Res.* **34**, 6044–6050 (2006).

42. Manley, K., Shirley, T. L., Flaherty, L. & Messer, A. *Msh2* deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genet.* **23**, 471–473 (1999).

43. Savouret, C. et al. CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.* **22**, 2264–2273 (2003).

44. van den Broek, W. J. et al. Somatic expansion behaviour of the (CTG)_n repeat in myotonic dystrophy knock-in mice is differentially affected by *Msh3* and *Msh6* mismatch-repair proteins. *Hum. Mol. Genet.* **11**, 191–198 (2002).

45. Kovtun, I. V. & McMurray, C. T. Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genet.* **27**, 407–411 (2001).

46. Owen, B. A. et al. (CAG)_n-hairpin DNA binds to *Msh2*–*Msh3* and changes properties of mismatch recognition. *Nature Struct. Mol. Biol.* **12**, 663–670 (2005).

47. Savouret, C. et al. *MSH2*-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol. Cell. Biol.* **24**, 629–637 (2004).

48. Yoon, S. R., Dubeau, L., de Young, M., Wexler, N. S. & Arnheim, N. Huntington disease expansion mutations in humans can occur before meiosis is completed. *Proc. Natl. Acad. Sci. USA* **100**, 8834–8838 (2003).

49. Anvret, M. et al. Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum. Mol. Genet.* **2**, 1397–1400 (1993).

50. Kennedy, L. et al. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.* **12**, 3359–3367 (2003).

51. Lia, A. S. et al. Somatic instability of the CTG repeat in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities. *Hum. Mol. Genet.* **7**, 1285–1291 (1998).

52. Fortune, M. T., Vassilopoulos, C., Coolbaugh, M. I., Siciliano, M. J. & Monckton, D. G. Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum. Mol. Genet.* **9**, 439–445 (2000).

53. Kovtun, I. V. et al. OGG1 initiates age-dependent CAG expansion in somatic cells during base excision repair of oxidized bases *in vitro* and *in vivo*. *Nature* **447**, 447–452 (2007).

54. Spiro, C. et al. Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol. Cell* **4**, 1079–1085 (1999).

55. Henriksen, L. A., Tom, S., Liu, Y. & Bambara, R. A. Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion. *J. Biol. Chem.* **275**, 16420–16427 (2000).

56. Panigrahi, G. B., Lau, R., Montgomery, S. E., Leonard, M. R. & Pearson, C. E. Slipped (CTG)•(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nature Struct. Mol. Biol.* **12**, 654–662 (2005).

57. Liu, Y., Zhang, H., Veeraraghavan, J., Bambara, R. A. & Freudenreich, C. H. *Saccharomyces cerevisiae* flap endonuclease 1 uses flap equilibration to maintain triplet repeat stability. *Mol. Cell. Biol.* **24**, 4049–4064 (2004).

58. van den Broek, W. J., Nelen, M. R., van der Heijden, G. W., Wansink, D. G. & Wieringa, B. Fen1 does not control somatic hypermutability of the (CTG)_n•(CAG)_n repeat in a knock-in mouse model for DM1. *FEBS Lett.* **580**, 5208–5214 (2006).

59. Warren, S. T. Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. *Science* **275**, 408–409 (1997).

60. Richards, R. I. et al. Evidence of founder chromosomes in fragile X syndrome. *Nature Genet.* **1**, 257–260 (1992).

61. Jakutciak, J. P. & Wells, R. D. Genetic instabilities in (CTG)•(CAG) repeats occur by recombination. *J. Biol. Chem.* **274**, 23468–23479 (1999).

62. Napierala, M., Dere, R., Vetcher, A. & Wells, R. D. Structure-dependent recombination hot spot activity of GAA•TTC sequences from intron 1 of the Friedreich's ataxia gene. *J. Biol. Chem.* **279**, 6444–6454 (2004).

63. Dere, R. & Wells, R. D. DM2 CCTG•CAGG repeats are crossover hotspots that are more prone to expansions than the DM1 CTG•CAG repeats in *Escherichia coli*. *J. Mol. Biol.* **360**, 21–36 (2006).

64. Freudenreich, C. H., Kantrow, S. M. & Zakian, V. A. Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279**, 853–856 (1998).

65. Nag, D. K., Suri, M. & Stenson, E. K. Both CAG repeats and inverted DNA repeats stimulate spontaneous unequal sister-chromatid exchange in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **32**, 5677–5684 (2004).

66. Richard, G.-F., Goellner, G. M., McMurray, C. T. & Haber, J. E. Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11-RAD50-XRS2 complex. *EMBO J.* **19**, 2381–2390 (2000).

67. Richard, G.-F., Cyncynatus, C. & Dujon, B. Contractions and expansions of CAG•CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a *MUS81*-independent mechanism. *J. Mol. Biol.* **326**, 769–782 (2003).

68. Meservy, J. L. et al. Long CTG tracts from the myotonic dystrophy gene induce deletions and rearrangements during recombination at the APRT locus in CHO cells. *Mol. Cell. Biol.* **23**, 3152–3162 (2003).

69. Bacolla, A., Wojciechowska, M., Kosmider, B., Larson, J. E. & Wells, R. D. The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA Repair (Amst.)* **5**, 1161–1170 (2006).

70. Rolfsmeier, M. L., Dixon, M. J. & Lahue, R. S. Mismatch repair blocks expansions of interrupted trinucleotide repeats in yeast. *Mol. Cell* **6**, 1501–1507 (2000).

71. Mirkin, S. M. & Smirnova, E. V. Positioned to expand. *Nature Genet.* **31**, 5–6 (2002).

72. Cleary, J. D. & Pearson, C. E. Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. *Trends Genet.* **21**, 272–280 (2005).

73. Abu-Baker, A. & Rouleau, G. A. in *Genetic Instabilities and Neurological Diseases* (eds Wells, R. D. & Ashizawa, T.) 487–513 (Elsevier, Amsterdam, 2006).

74. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H. & Housman, D. E. Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl. Acad. Sci. USA* **94**, 7388–7393 (1997).

75. Amack, J. D. & Mahadevan, M. S. The myotonic dystrophy expanded CUG repeat tract is necessary but not sufficient to disrupt C2C12 myoblast differentiation. *Hum. Mol. Genet.* **10**, 1879–1887 (2001).

76. Mankodi, A. et al. Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* **289**, 1769–1772 (2000).

77. Tassone, F. et al. Elevated levels of *FMR1* mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.* **66**, 6–15 (2000).

78. Jin, P. et al. RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in *Drosophila*. *Neuron* **39**, 739–747 (2003).

79. Mutusdi, M., Marshall, C. M., Benzow, K. A., Koob, M. D. & Rebay, I. The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in *Drosophila*. *Curr. Biol.* **14**, 302–308 (2004).

80. Lin, X. & Ashizawa, T. Recent progress in spinocerebellar ataxia type-10 (SCA10). *Cerebellum* **4**, 37–42 (2005).

81. Napierala, M. & Krzyzosiak, W. J. CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J. Biol. Chem.* **272**, 31079–31085 (1997).

82. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. & Krzyzosiak, W. J. RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.* **31**, 5469–5482 (2003).

83. Handa, V., Yeh, H. J., McPhie, P. & Usdin, K. The AUUCU repeats responsible for spinocerebellar ataxia type 10 form unusual RNA hairpins. *J. Biol. Chem.* **280**, 29340–29345 (2005).

84. Napierala, M., Michalowski, D., de Mezer, M. & Krzyzosiak, W. J. Facile *FMR1* mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res.* **33**, 451–463 (2005).

85. Fardaei, M. et al. Three proteins, MBNL, MBLL and MBXL, co-localize *in vivo* with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* **11**, 805–814 (2002).

86. Miller, J. W. et al. Recruitment of human muscleblind proteins to (CUG) expansions associated with myotonic dystrophy. *EMBO J.* **19**, 4439–4448 (2000).

87. Lu, X., Timchenko, N. A. & Timchenko, L. T. Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. *Hum. Mol. Genet.* **8**, 53–60 (1999).

88. Philips, A. V., Timchenko, L. T. & Cooper, T. A. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**, 737–741 (1998).

89. Thornton, C. A., Swanson, M. S. & Cooper, T. A. in *Genetic Instabilities and Neurological Diseases* (eds Wells, R. D. & Ashizawa, T.) 37–54 (Elsevier, Amsterdam, 2006).

90. Ho, T. H. et al. Muscleblind proteins regulate alternative splicing. *EMBO J.* **23**, 3103–3112 (2004).

91. Kino, Y. et al. Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.* **13**, 495–507 (2004).

92. Kanadia, R. N. et al. A muscleblind knockout model for myotonic dystrophy. *Science* **302**, 1978–1980 (2003).

93. Kanadia, R. N. et al. Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proc. Natl. Acad. Sci. USA* **103**, 11748–11753 (2006).

94. Iwahashi, C. K. et al. Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**, 256–271 (2006).

95. Malinina, L. Possible involvement of the RNAi pathway in trinucleotide repeat expansion diseases. *J. Biomol. Struct. Dyn.* **23**, 233–235 (2005).

96. Handa, V., Saha, T. & Usdin, K. The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. *Nucleic Acids Res.* **31**, 6243–6248 (2003).

97. Krol, J. et al. Ribonuclease Dicer cleaves triplet repeat hairpins into shorter repeats which silence specific targets. *Mol. Cell* **25**, 575–586 (2007).

98. Cho, D. H. et al. Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* **20**, 483–489 (2005).

99. Moseley, M. L. et al. Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nature Genet.* **38**, 758–769 (2006).

100. Gray, S. J., Gerhardt, J., Doerfler, W., Small, L. E. & Fanning, E. An origin of DNA replication in the promoter region of the human fragile X mental retardation (*FMR1*) gene. *Mol. Cell. Biol.* **27**, 426–437 (2007).

Acknowledgements I thank W. Krzyzosiak, R. Lahue, K. Lobachev, C. McMurray, D. Monckton, C. Pearson, M. Swanson, K. Usdin and R. Wells for sharing their ideas and unpublished results. I also extend my gratitude to all the participants of the 5th International Conference on Unstable Microsatellites and Human Disease (Granada, Spain, 2006) for their intense and productive discussions, which helped to shape this review. I am indebted to my wife, Kate, for her invaluable critical comments. I thank J. White and P. White for their generous support. This work was supported by the National Institutes of Health.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (sergei.mirkin@tufts.edu).

Base-excision repair of oxidative DNA damage

Sheila S. David¹, Valerie L. O'Shea² & Sucharita Kundu²

Maintaining the chemical integrity of DNA in the face of assault by oxidizing agents is a constant challenge for living organisms. Base-excision repair has an important role in preventing mutations associated with a common product of oxidative damage to DNA, 8-oxoguanine. Recent structural studies have shown that 8-oxoguanine DNA glycosylases use an intricate series of steps to locate and excise 8-oxoguanine lesions efficiently against a high background of undamaged bases. The importance of preventing mutations associated with 8-oxoguanine is shown by a direct association between defects in the DNA glycosylase MUTYH and colorectal cancer. The properties of other guanine oxidation products and the associated DNA glycosylases that remove them are now also being revealed.

The specific pairing of DNA bases — A with T and G with C — is crucial for preserving the information content of the genome. However, the structural properties of DNA bases, and therefore their pairing properties, are often modified by reactions with environmental toxins (for example, ultraviolet radiation and cigarette smoke) and endogenous metabolic products or by-products (for example, S-adenosylmethionine and hydroxyl radicals)^{1–3}. Over three decades ago, armed with the knowledge that cytosine was readily deaminated to yield uracil, T. Lindahl searched for bacterial enzymes that mediated the repair of uracil and discovered uracil DNA glycosylase (as discussed in ref. 4). This enzyme catalyses the removal of uracil from DNA, the first step of the base-excision repair (BER) pathway^{5–7}. We now know that there are many DNA glycosylases that specifically recognize and excise abnormal DNA bases. After the base has been removed, other enzymes are recruited to excise the remaining sugar fragments and to reinstall an undamaged nucleotide, thereby repairing the damaged base site⁸.

Recent studies have propelled the DNA-damage-specific glycosylases into the limelight. Structural, biophysical and biochemical approaches have provided exquisite insight into the remarkable process by which these enzymes locate damaged bases in the context of a large excess of normal DNA. Genetic studies have revealed direct correlations between defects in the repair of damaged bases and human disease. And sophisticated analytical and chemical approaches have identified new base lesions and their associated glycosylases. Here, we highlight the BER of guanine oxidation products (Fig. 1) and illustrate notable features of the chemistry, structural biology and clinical aspects of these repair pathways.

The special problem of 8-oxoguanine

The production of reactive oxygen species (ROS), such as hydrogen peroxide, superoxide and hydroxyl radicals, has been linked to the initiation and progression of cancer⁹. ROS are by-products of cellular respiration and components of inflammatory responses. In addition, ROS can be produced as a consequence of ionizing radiation or environmental exposure to transition metals, chemical oxidants and free radicals. Normal cells have enzymatic and non-enzymatic mechanisms to counter the production of ROS. But aberrantly functioning cells are often in a state of 'oxidative stress', in which the balance between

oxidants and antioxidants has been disrupted, resulting in increased levels of cellular damage.

DNA bases are particularly susceptible to oxidation mediated by ROS¹⁰. The low redox potential of guanine makes this base particularly vulnerable and leads to a plethora of oxidized guanine products¹⁰. The most thoroughly examined guanine oxidation product is 7,8-dihydro-8-oxoguanine (also known as 8-oxoguanine; 8-oxoG)^{10,11}; the presence of 8-oxoG is often used as a cellular biomarker to indicate the extent of oxidative stress⁹. The 8-oxoG lesion is particularly deleterious because of its two-atom change from G (the introduction of an oxo group on the carbon at position 8 (C8) and an hydrogen atom on the nitrogen at position 7 (N7)) and because of its ability to mimic T functionally in the *syn* conformation, forming a stable 8-oxoG(*syn*)•A(*anti*) base pair (Fig. 2a). In contrast to many other types of DNA damage, these structural features allow efficient, although inaccurate, bypass of 8-oxoG by replicative DNA polymerases¹². A series of X-ray crystal structures of a DNA polymerase I fragment (BF) from *Bacillus stearothermophilus* (also known as *Geobacillus stearothermophilus*) illustrated the consequences of 8-oxoG presence in the template before insertion of dAMP or dCMP, after insertion, and after extension beyond 8-oxoG•A and 8-oxoG•C base pairs¹³. A comparison of these structures with those containing DNA mismatches showed that inaccurate bypass of 8-oxoG results from the ability of the 8-oxoG(*syn*)•A(*anti*) base pair to mimic a normal base pair when processed by BF. By contrast, the formation of an 8-oxoG(*anti*)•C(*anti*) base pair during replication induces template and polymerase distortions similar to those seen when the active site of DNA polymerase encounters mismatches. The structural data provide a molecular rationale for both the more facile insertion of A than C opposite 8-oxoG and the lack of error detection by replicative DNA polymerases after this insertion.

Failure to remove 8-oxoG before replication results in G-to-T transversion mutations¹⁰. Repair pathways in all organisms are tailored to deal with the special features of 8-oxoG and to mitigate its high mutagenic potential. The MutT, MutM (also known as Fpg) and MutY enzymes in bacteria¹⁴ and the corresponding human enzymes⁷, MTH1, OGG1 and MUTYH (formerly hMYH), constitute the 8-oxoG repair pathway (also known as the GO repair pathway) (Fig. 1). MutT (or MTH1) hydrolyses 8-oxo-dGTP, removing it from the nucleotide

¹Department of Chemistry, University of California at Davis, 1 Shields Avenue, Davis, California 95616, USA. ²Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112, USA.

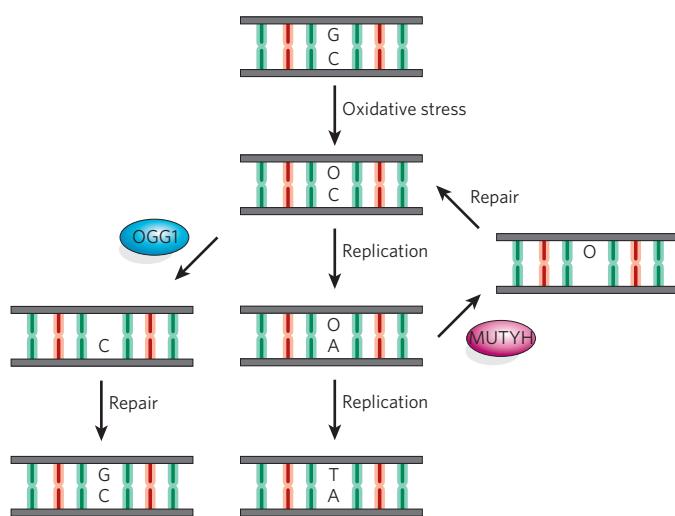


Figure 1 | Short-patch BER pathway for 8-oxoG. The presence of 8-oxoG (O) in DNA causes G-to-T transversions, as illustrated in the central pathway. The human DNA glycosylases OGG1 and MUTYH are involved in excision of bases from the DNA. OGG1 removes 8-oxoG from 8-oxoG•C base pairs, and MUTYH removes A from 8-oxoG•A base pairs, both generating AP sites in the DNA. The corresponding bacterial enzymes are MutM and MutY. The steps labelled ‘repair’ summarize the actions of AP endonuclease, deoxyribophosphate lyase, DNA polymerase and DNA ligase. Notably, OGG1 also has AP lyase activity. MutT and its human homologue MTH1 (not shown) have an important role in preventing the incorporation of 8-oxoG, through hydrolysis of free 8-oxo-dGTP.

pool so that it cannot be incorporated into DNA by polymerases. The BER glycosylase MutM (OGG1) excises 8-oxoG from the 8-oxoG•C base pair so that subsequent processing by other enzymes in the BER pathway can restore the G•C base pair. However, if this does not occur, and replication takes place, then the DNA glycosylase MutY (MUTYH) intercepts the resultant 8-oxoG•A base pair and removes the inappropriate A. Subsequent processing of the apurinic/apyrimidinic (AP) site, followed by replication by a repair polymerase, provides an opportunity to create an 8-oxoG•C substrate for the 8-oxoG DNA glycosylase, MutM (OGG1). Notably, structural studies of the eukaryotic repair polymerase- β with 8-oxoG-containing templates have shown that the active site of this polymerase more easily accommodates 8-oxoG•C base pairs than 8-oxoG•A mismatches¹⁵. This specificity is exactly the opposite of the results obtained with the bacterial polymerase BF. Moreover, this structural information explains the higher propensity of polymerase- β to insert dCTP over dATP opposite 8-oxoG, a property that is required of this repair polymerase to prevent futile cycles of repair and replication¹⁵.

Seeking and removing 8-oxoG lesions

Recognition of 8-oxoG lesions by OGG1

In eukaryotic nuclear DNA, it has been estimated that there are several 8-oxoG bases per 10^6 guanine bases¹⁶. This low frequency and the similarity in structure of 8-oxoG and G make it a formidable challenge for 8-oxoG DNA glycosylases to find 8-oxoG lesions among a huge number of undamaged bases. Other BER glycosylases, such as uracil DNA glycosylase and MutY (MUTYH), also need to distinguish relatively minor perturbations from the normal DNA bases or base pairs. Structural and biophysical studies of many DNA glycosylases bound to their relevant substrate duplexes have revealed common themes in lesion recognition, including enzyme-initiated DNA bending, disruption of the base pair, extrusion of the damaged nucleotide from the interior of the DNA helix, and placement of the target base into a base-specific enzyme pocket^{17–21}. This enzyme-driven process is often referred to as ‘base flipping’ but might be described more accurately as ‘nucleotide

flipping’, because the entire nucleotide is rotated out of the helix to accommodate the base in the base-specific pocket²⁰. The molecular ‘gymnastics’ mediated by the DNA-damage-specific glycosylases have been described as a routine involving specific ‘pinch-push-plug-pull’ steps^{17,18}. On encountering a specific DNA lesion, the DNA-damage-specific glycosylase induces bending and distortion of the DNA double helix, constituting the ‘pinch’. This step, as well as a ‘push’ of the target base out of the helix, is facilitated by intercalation of a glycosylase amino-acid side chain into the DNA helix. The same or another intercalating amino acid functions as a ‘plug’ to fill the position of the extruded base and stabilize the contortion of the DNA duplex. A ‘pull’ by active-site residues specific for the relevant target base secures it into the recognition and clipping pocket.

These general features used for lesion recognition are well illustrated by the X-ray crystal structure of a catalytically inactive variant of OGG1 (in which the lysine residue at position 249 is replaced by a glutamine residue; Lys249Gln) bound to a substrate duplex containing an 8-oxoG•C base pair²². This variant of OGG1 retains high-affinity binding to an 8-oxoG•C-containing duplex but cannot excise 8-oxoG because of the mutation of Lys 249, which participates in catalysis of the base-displacement reaction. Such enzyme complexes with a lesion-containing DNA substrate are referred to as lesion-recognition complexes (LRCs). In the X-ray crystal structure of the LRC described above, the 8-oxoG•C base pair has been disrupted, and the 8-oxoG nucleotide has been rotated

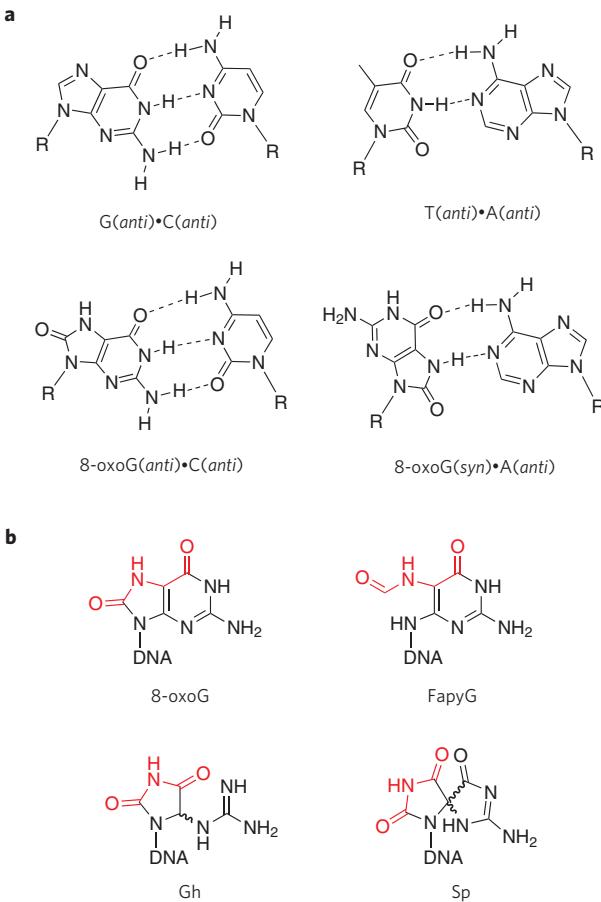


Figure 2 | Structures of 8-oxoG-containing base pairs and of several nucleosides of guanine oxidation products. **a**, The structures of the base pairs G•C and T•A are compared with those of 8-oxoG•C and 8-oxoG•A. 8-oxoG differs from G by an oxo group at C8 and an NH at N7. This subtle change allows 8-oxoG to base-pair easily with either C or A. **b**, A large number of guanine oxidation products have been observed¹⁰. Here, we focus on the hydantoin lesions: Gh and Sp. The structures of these lesions resemble 8-oxoG and FapyG in retaining hydrogen-bonding functionality (red) that can mimic T.

to place the attached base within an 8-oxoG-specific pocket (Figs 3 and 4a). Although OGG1 makes many specific contacts with both the extrahelical 8-oxoG and the orphaned C, only one hydrogen-bonding contact would not be possible with G. This crucial hydrogen bond is between the carbonyl oxygen of Gly 42 of OGG1 and the hydrogen at N7 of 8-oxoG. With G instead of 8-oxoG, not only would such a hydrogen bond be absent, but there would also be unfavourable repulsion between the carbonyl oxygen and the lone pair at N7. It seems unlikely that specificity for 8-oxoG over G would be due to a single hydrogen bond, so the mechanism for discrimination between these two bases is probably much more complex.

Revealing encounters of OGG1 with G•C base pairs

Understanding the strategy used by OGG1 to discriminate between 8-oxoG•C and G•C required a structural glimpse of OGG1 encountering a G•C base pair. Because these enzymes do not recognize undamaged G•C base pairs, such an encounter is necessarily transient. For this reason, Verdine and co-workers used an innovative covalent trapping method to secure OGG1 in proximity to a G•C base pair²³. Using the X-ray crystal structure of the OGG1 LRC as a guide²², a cysteine residue was installed at position 149 of the enzyme (in place of asparagine), and a single thiol-modified-base-containing nucleotide was positioned within the DNA duplex such that a disulphide crosslink would be formed when OGG1 encountered the targeted G•C base pair. The remarkable structure²³ revealed that even though G is forcibly presented to the enzyme, it does not gain access to the 8-oxoG site but is lodged within an alternative, peripheral, 'exo' site (Fig. 4b). Calculations of differences in free energy indicate that favourable 8-oxoG interactions and unfavourable G interactions in both the active and exo sites result in the 10⁵-fold preference for 8-oxoG over G. Quantum calculations also show that the 8-oxoG and G bases have local dipoles that have opposite orientations as a result of the charge inversion at C8 and N7. The active site of OGG1 has a complementary dipole for 8-oxoG created by a Lys-249-NH₃⁺ and Cys-253-S⁻ amino-acid pair within the active site (Fig. 3). Indeed, the contribution of this dipole–dipole interaction to recognition is calculated to be greater than the contribution of the single hydrogen bond to NH7 of 8-oxoG. However, the importance of the specific hydrogen bond to 8-oxoG was shown by replacing Gly 42 with alanine and carrying out structural analysis of the resultant complexes using the disulphide-crosslink approach with 8-oxoG or G presented to the enzyme²⁴. Surprisingly, in all of the structures examined, the backbone conformation at position 42 remains identical to that in the LRC (Fig. 3), even though the alanine substitution is energetically unfavourable as a result of the steric demands of the side-chain methyl group. These structures show the role of the surrounding protein in 'hard-wiring' the conformation of Gly 42 to preserve the ability of OGG1 to recognize 8-oxoG and effectively reject G from the active site²⁴.

The presence of the exo base-binding site, and the calculations showing preferences for binding to 8-oxoG at both base-binding sites, suggest that the damaged base is extracted from the duplex by OGG1 in discrete steps: an initial step involves placement into the exo site, and a final step involves engagement in the active-site pocket. The use of a multistep base-excision process allows many opportunities to check the identity of the base before its excision. In the case of OGG1, only 8-oxoG has the proper 'credentials' to make it through all of the checkpoints.

Additional intermediates in the 8-oxoG repair pathway

A later intermediate in the 8-oxoG recognition and excision process was revealed recently by structural analysis of several constructs containing a mutation (Gln315Phe) in OGG1 that enlarges a side chain to occlude the active-site pocket²⁴ (Fig. 3). Notably, in one structure, the 8-oxoG lesion is almost completely inserted into the active-site pocket; however, the 8-oxoG base is not clipped out of the DNA by this mutated enzyme. This demonstrates another level of quality control, requiring perfect geometric alignment of the base within the specificity pocket to attain the proper transition state for base cleavage.

The disulphide-crosslink approach also provided a fortuitous structure of OGG1 'interrogating' a G•C base pair located adjacent to an 8-oxoG•C base pair in a DNA duplex²⁵ (Fig. 4c). This structure reveals how 8-oxoG might first be detected by OGG1 during the evaluation of the neighbouring base pair. The overall structure is similar to the LRCs and the interrogation complex with the G•C base pair discussed earlier. The target G, however, is destacked from the helix, but it does not reside at the exo site. Instead, the target G is folded back into the major groove and forms an unusual triple hydrogen-bonding interaction with the major groove face of the neighbouring 8-oxoG. The altered mode of examination of G in the vicinity of 8-oxoG seems to be associated with severely repulsive steric and electrostatic 'clashes' with the phosphate backbone as a result of the presence of the 8-oxo group of 8-oxoG. The altered backbone conformations do not allow complete presentation of the G to the exo site, thus favouring the folded-in conformation that is further stabilized by hydrogen bonding with 8-oxoG. The authors of this study, Banerjee and Verdine, noted that such base–phosphate clashes only occur as a result of the remodelling of the duplex by OGG1 (ref. 25). Thus, the active role of the DNA glycosylase in probing and altering the duplex is a key feature of the mechanism for detecting subtle alterations in the base structure.

Intrahelical interrogation of normal base pairs

The structures described here indicate that G•C base pairs might be paid special attention by OGG1. But does OGG1 scrutinize normal base pairs by removing one of the bases from the helix? Such a process would seem to be energetically costly and inefficient. To explore further the interrogation process used by 8-oxoG DNA glycosylases, the disulphide-crosslink approach was used to generate several structures of *B. stearothermophilus* MutM (which is functionally similar to OGG1) tethered to normal DNA in proximity to both an A•T and a G•C base pair²⁶ (Fig. 4d–f). In these structures, the normal bases are not extruded from the helix, suggesting that MutM initially locates 8-oxoG lesions by intrahelical interrogation. Indeed, the side chain of a phenylalanine residue (Phe 114) was found to be wedged into the helix from the minor groove above the A•T or G•C base pair. This invasion by Phe 114 results in severe bending of the DNA and buckling of the base pair. In addition, two other amino-acid residues (Arg 112 and Met 77) that take the place

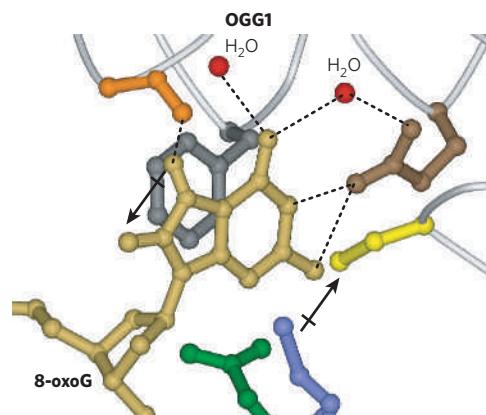


Figure 3 | Recognition of 8-oxoG by OGG1 observed in the LRC of OGG1 with 8-oxoG-C-containing duplexes. This is a view of the base-specific pocket of OGG1, showing residues involved in the recognition of 8-oxoG. DNA is shown in gold, with the OGG1 backbone in light grey and coordinating water molecules in red. Residues that are important for recognition are shown as follows: orange (Gly 42), purple (Lys 249), yellow (Cys 253), green (Asp 268), brown (Gln 315) and dark grey (Phe 319). Hydrogen bonds are shown as dashed lines. An 8-oxoG-specific hydrogen bond between NH7 and the carbonyl of Gly 42 provides specificity for 8-oxoG over G. In addition, the dipole (crossed arrow) that is associated with the Lys 249–Cys 253 pair makes favourable dipole–dipole interactions with 8-oxoG but not with G. Image generated from pdb file 1EBM from the Worldwide Protein Data Bank, based on data from ref. 22.

of the flipped-out 8-oxoG base in the corresponding LRC²⁷ are drawn back, as though they are waiting to lunge in when space has been created. The overall structure of these non-specific complexes is similar to that of the LRC, except that many contacts between the enzyme and the DNA are water-mediated rather than direct. This might allow a looser association with the DNA, facilitating movement along the helix. Interestingly, despite using the disulphide-crosslink strategy, it was difficult to trap a complex in which MutM was interrogating a G•C base pair; indeed, the protein seemed to avoid interrogating G•C base pairs. To obtain the desired structure, the DNA sequence was altered so that a string solely consisting of G•C base pairs was presented to the enzyme. Taken together, the various structures indicate that Phe 114 is a sensor of the stability and/or deformability of the target base pair. The authors, Banerjee *et al.*, suggest that although an 8-oxoG•C base pair is only slightly less stable than an A•T base pair, the non-standard properties of this base pair might render it unable to withstand the insertion of Phe 114 of MutM²⁶. After the base pair has been disrupted, the liberated 8-oxoG would then be captured by the active site of MutM.

A redundant search of the genome for 8-oxoG lesions

It might be expected that such a search process would not be extremely precise; therefore, Banerjee *et al.* suggest that DNA glycosylases might compensate for this imprecision by being extremely fast, allowing

repeated opportunities to find a damaged base²⁶. This fast search process was visualized recently using single-molecule detection to track OGG1 movements along a normal DNA duplex²⁸. In these studies, OGG1 was found to move along the DNA with a diffusion constant approaching the theoretical upper limit for one-dimensional diffusion, indicating that OGG1 samples millions of base pairs per second. On the basis of these measurements, the estimated barrier to sliding is extremely small ($0.5 \text{ kcal mol}^{-1}$). The smaller barrier and the observed unbiased random movement of OGG1 on DNA suggest that OGG1 rapidly searches along DNA as a consequence of brownian motion. Effects on the measured diffusion constants as a result of mutation of His 270, which is located at the DNA-binding interface^{22–24}, and pH suggest that the DNA–protein interface has been optimized to allow fast enzyme sliding.

Taken together, a scheme of the events involved in the search for 8-oxoG lesions by OGG1 or MutM can be proposed from the results of these studies (Fig. 5). First, the 8-oxoG DNA glycosylase moves rapidly along the helix, inserting the probe ligand (for example, Phe 114 in MutM) into the helix to look for vulnerable sites. Intercalation of an amino-acid residue of the enzyme at a normal base pair merely buckles the base pair; however, it can be envisaged that such a probing event could disrupt an abnormal base pair such as 8-oxoG•C. Such a search process would be extremely fast, so an 8-oxoG base might be missed

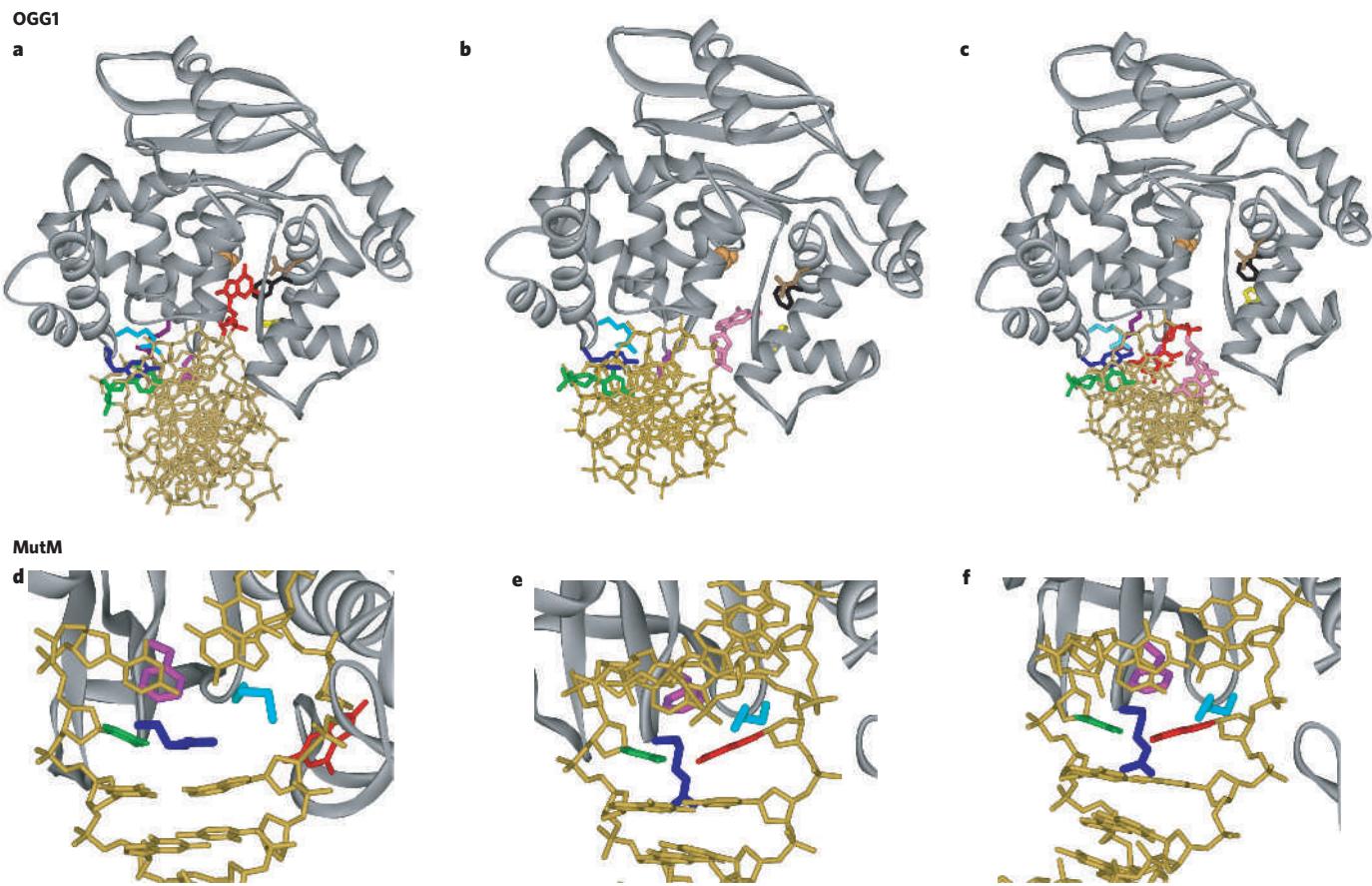


Figure 4 | The LRCs of OGG1 and MutM with non-specific complexes

(normal base pairs). In all structures, the protein backbone is shown as a dark grey ribbon, and the DNA is in gold. In **a–c**, important amino-acid side chains are shown in orange (Gly 42), dark pink (Asn 149 or Cys 149), light purple (Arg 154), dark purple (Tyr 203), light blue (Arg 204), yellow (His 270), brown (Gln 315) and black (Phe 319). **a**, OGG1 LRC with 8-oxoG•C-containing DNA. 8-oxoG is shown in red, and the estranged C in green. **b**, OGG1 interrogating a G•C base pair. The target G is shown in light pink, and the target C in green. (Arg 154 is not shown.) **c**, OGG1 interrogating a G•C base pair adjacent to an 8-oxoG lesion. 8-oxoG is

shown in red, the target G in light pink, and the target C in green. In **d–f**, the intercalating residue, Phe 114, is shown in dark pink, and Met 77 and Arg 112 are in light blue and dark purple, respectively. **d**, MutM LRC with 8-oxoG•C-containing DNA. 8-oxoG is shown in red, and the estranged C in green. **e**, MutM interrogating a G•C base pair. The target G is shown in red, and the target C in green. **f**, MutM interrogating an A•T base pair. The target A is shown in red, and the target T in green. Images generated from pdb files from the Worldwide Protein Data Bank, based on data from the following: ref. 22, file 1EBM (**a**), ref. 23, file 1YQK (**b**), ref. 25, file 2I5W (**c**), ref. 27, file 1R2Y (**d**), ref. 26, file 2F5O (**e**) and ref. 26, file 2F5N (**f**).

sometimes. But this might be preferable to wasting time interrogating undamaged DNA. This fast, random and redundant process would provide many opportunities to find 8-oxoG. After the 8-oxoG has been partly expelled from the helix, it is captured by the exo site and passed quickly to the 8-oxoG-specific active site. Occasionally, a G might be displaced from the helix; however, it would be intercepted in the exo base-binding site and would not proceed to the 8-oxoG-specific pocket; instead, it would be placed back in the helix. Moreover, if a G adjacent to 8-oxoG is interrogated, it does not fully dislodge from the helix, thus allowing the 8-oxoG DNA glycosylase to home in more rapidly on the adjacent 8-oxoG lesion. Structural and biophysical studies suggest other DNA glycosylases probably operate by similar mechanisms^{18,29}. Indeed, some damaged bases might be expelled more readily from the helix or be distinguished more readily from their normal counterparts, making such a search process more streamlined.

Recognition of 8-oxoG•A mismatches by MutY and MUTYH

The partners for MutM and OGG1 in the 8-oxoG repair pathway, MutY and MUTYH, also need to recognize 8-oxoG precisely, for selective removal of only the A bases present in 8-oxoG•A mismatches. MutY has a special carboxy-terminal domain that is not found in other BER glycosylases, and this domain has an important role in the recognition of 8-oxoG^{30–32}. In the X-ray crystal structure of an inactive variant (Asp144Asn) of *B. stearothermophilus* MutY³², there are a plethora of contacts with 8-oxoG but minimal contacts with A. Despite the implied importance of 8-oxoG, MutY also mediates the removal of A opposite other bases, including G, FapyG (a formamidopyrimidine) and C^{5,33}. Time-resolved fluorescence experiments of the MutY A-excision reaction using 8-oxoG•A substrates indicated a multiphase reaction profile, with a fast process being associated with changes at 8-oxoG and a slower process associated with altering the environment of the A³⁴. The fluorescence data were originally interpreted as indicating sequential extrusion of 8-oxoG then A from the helix. However, in the *B. stearothermophilus* MutY LRC structure³², 8-oxoG is within the helix but has an altered conformation about the N-glycosidic bond. Although the molecular mechanism of the 8-oxoG-associated fluorescence changes is unclear, these experiments, together with the structure, suggest that MutY relies heavily on initial recognition of 8-oxoG to locate A bases for excision. Additional structures of MutY bound to other base pairs (such as T•A) would be of considerable interest to elaborate how this enzyme prevents inadvertent excision of A opposite T.

BER and colorectal cancer

The importance of preventing mutations associated with 8-oxoG was emphasized when a direct link was uncovered between colorectal cancer and mutations in the gene encoding the human MutY homologue (MUTYH)³⁵. This work also established the first link between inherited defects in BER and cancer. In a British family (denoted family N), several siblings presented clinical symptoms characteristic of familial adenomatous polyposis, a common form of familial colon cancer. In familial adenomatous polyposis, the colon of afflicted individuals is littered with adenomatous polyps as a result of mutations in the gene adenomatous polyposis coli (*APC*)³⁶. The APC protein has many important roles in controlling the proliferation of colon cells and is mutated in most colorectal tumours. Although the familial nature of familial adenomatous polyposis is usually a consequence of inherited mutations in *APC*, this was not the case in family N. Consistent with the polyposis phenotype, DNA from tumours of afflicted members of family N had inactivating mutations in *APC*. The types of mutation, G-to-T transversions, was particularly revealing, because this type of mutation is commonly associated with 8-oxoG and therefore prompted sequencing of the genes encoding the enzymes of the human 8-oxoG repair pathway. This revealed germline biallelic missense mutations in *MUTYH* that would result in two variants of MUTYH, each containing a single amino-acid substitution, Tyr165Cys and Gly382Asp. Owing to the high similarity of MUTYH and *Escherichia coli* MutY,

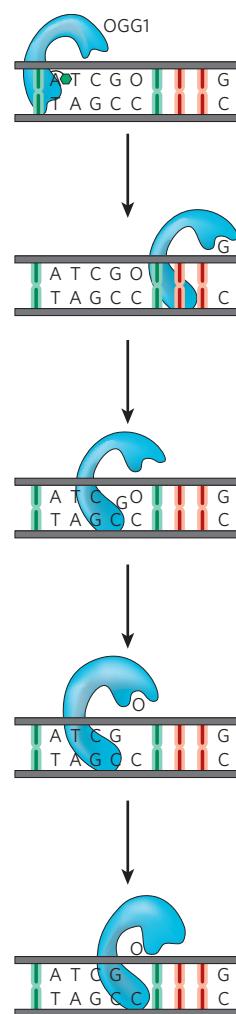


Figure 5 | The 8-oxoG lesion search process. The pathway is based on structures of OGG1 and MutM ‘trapped’ with DNA. The enzyme tracks rapidly along DNA, inserting a ‘probe’ amino-acid residue (green hexagon) at various base pairs to test the stability and/or deformability of the duplex. This results in preferential expulsion of 8-oxoG (O) from 8-oxoG•C base pairs. However, if a G from a G•C base pair is extruded, it is captured by the exo site and then replaced within the helix. Encountering a G•C base pair adjacent to an 8-oxoG•C base pair enables 8-oxoG to be detected easily, because this G cannot be extruded in the same manner, promoting movement of the enzyme to the 8-oxoG•C base pair. At this point, the 8-oxoG might first be extruded to the exo site and then be quickly captured in the 8-oxoG-specific pocket, where it is excised from the DNA.

the activities of the two corresponding variants in *E. coli* MutY were analysed³⁵. Both were found to have compromised A removal activity relative to wild-type MutY^{35,37}. This finding supported the hypothesis that the MUTYH variants have a reduced capacity to initiate the repair of 8-oxoG•A mismatches, leading to increased numbers of G-to-T transversions in *APC* and eventually resulting in inactivation of the APC protein. This is a novel mechanism by which inherited defects in a gene encoding a BER enzyme (*MUTYH*) lead to mutations in a gene associated with predisposition to cancer (*APC*). This new mechanism of predisposition to colorectal cancer is now referred to as MUTYH-associated polyposis^{38,39}.

To evaluate the consequences of the two amino-acid variants on DNA-damage and -mismatch recognition independently of base cleavage, dissociation constants of the two variant MutY enzymes were measured with DNA duplexes containing a non-cleavable A analogue, 2'-deoxy-2'-fluoroadenosine (FA), opposite 8-oxoG and G³⁷. In these experiments, the two MutY enzymes (with the substitutions Tyr82Cys and Gly253Asp) showed a lower affinity for both 8-oxoG•FA-containing duplexes and G•FA-containing duplexes than did the wild-type enzyme. Surprisingly, the affinity of the variants for the duplex was not increased by the presence of 8-oxoG, in contrast to the case for the wild-type enzyme, indicating that the mutations abrogate the ability to recognize 8-oxoG over G. Insight into the role of these two amino acids in 8-oxoG recognition was also provided by determining the structure of a *B. stearothermophilus* MutY variant that has an Asp144Asn substitution bound to an 8-oxoG•A-containing duplex³² (Fig. 6a). The corresponding tyrosine residue in *B. stearothermophilus* MutY (Tyr 88) was found to be intercalated 5' to 8-oxoG and to participate in a hydrogen-bond

network with NH7 of 8-oxoG, whereas the glycine residue (Gly 260) is located in a turn region of the protein where backbone amides are involved in close hydrogen bonds with the phosphodiester backbone adjacent to 8-oxoG.

Interestingly, the location of Tyr 82 in MutY corresponds to the Phe 114 'probe' amino acid revealed in the structure of MutM trapped with non-specific DNA²⁶. This suggests that the ability of MutY to seek and find 8-oxoG•A mismatches is hampered by mutation of the large tyrosine probe ligand to a smaller residue, cysteine. The ability to replace the tyrosine residue in MutY with a leucine residue without sacrificing catalytic activity established the importance of intercalation of a bulky residue for indirect recognition of 8-oxoG•A mismatches⁴⁰, because a leucine side chain would be unable to participate in the hydrogen-bond network with NH7 of 8-oxoG. Many DNA glycosylases have a phenylalanine, leucine or tyrosine residue at the analogous position²⁶. In a more subtle manner, the turn region of *B. stearothermophilus* MutY that contains Gly 260 seems to be stabilizing a pinched and distorted conformation of the phosphodiester backbone adjacent to 8-oxoG, and this might be required for promoting A extrusion. This glycine residue is located in the C-terminal domain of MutY, which has been shown to be important for 8-oxoG recognition^{30–32}. Interestingly, the single amino-acid change in the Gly253Asp MutY variant resulted in a loss in affinity for duplexes containing 8-oxoG over G that was similar to that observed with a truncated form of MutY completely lacking the

C-terminal domain³⁷. Study of a structure of the *B. stearothermophilus* Asp144Asn MutY variant bound to 8-oxoG•A-mismatch-containing DNA³² revealed that the peptide-backbone torsion angles of Gly 260 are highly unusual, and glycine is the only amino acid that would easily accommodate such a conformation. Thus, replacement of this glycine with aspartic acid or alanine destabilizes the turn region and alters the ability to recognize 8-oxoG at this site specifically⁴⁰.

Although both variant MutY enzymes have a DNA-binding defect, the binding defect of the Tyr82Cys variant *in vitro* translates into a more deleterious effect on A excision^{37,40}. The *in vitro* adenine glycosylase activity of Tyr165Cys and Gly382Asp MUTYH variants (A. L. Livingston and S.S.D., unpublished observations) and the corresponding variants of mouse MUTYH⁴¹ (also known as MYH) were analogous to those observed for the *E. coli* variants. However, in contrast to the *E. coli* variants, the activity of the mouse MUTYH variants is further reduced in the presence of other DNA-binding proteins^{41,42}. For example, the presence of the human AP endonuclease (APE1; also known as APEX1) increases the A-excision activity of wild-type mouse MUTYH by stimulating release of MUTYH from the AP-site DNA product. By contrast, for the mouse MUTYH variants, the presence of APE1 reduces the efficiency of A-excision activity⁴¹. This is a consequence of competition between MUTYH and APE1 for the 8-oxoG•A-containing DNA substrate, which is not a problem for the wild-type enzyme, owing to its greater affinity for 8-oxoG•A. These competition effects might also explain why, in

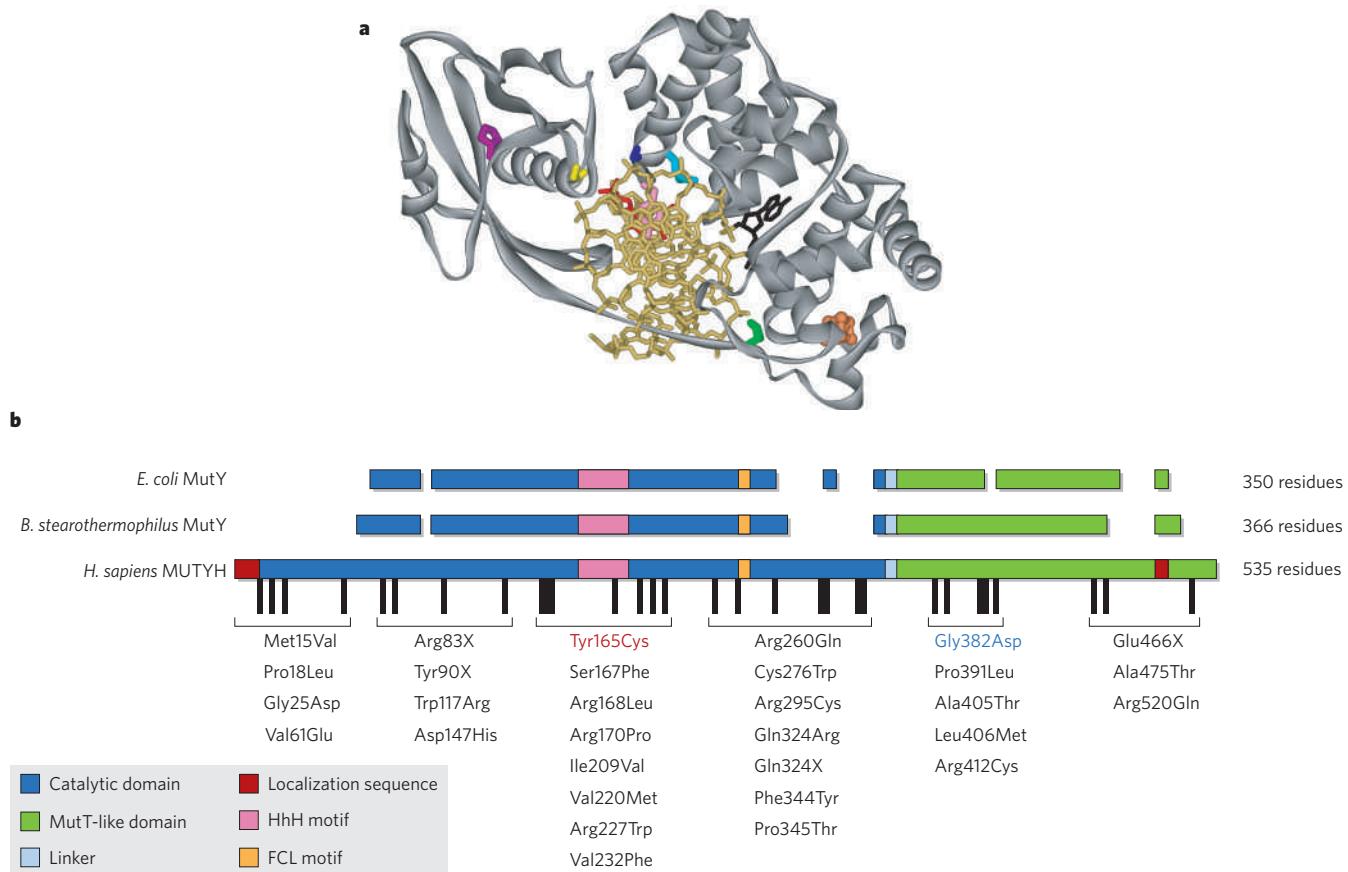


Figure 6 | Germline mutations observed in MUTYH in individuals with MUTYH-associated polyposis. **a**, Structure of *B. stearothermophilus* MutY with an Asp144Asn substitution bound to an 8-oxoG•A-mismatch-containing duplex. The view is down the helical axis, with the MutY backbone (grey ribbon) encircling the DNA (gold). The A is shown in black, bound in its recognition pocket, and 8-oxoG is shown in red. The 4Fe–4S cluster is shown in orange. Amino-acid residues in MutY that correspond to some of the residues observed to vary in individuals with MUTYH-associated polyposis are indicated on the MutY ribbon: MutY Tyr 88 (light pink), MUTYH Tyr165Cys; Ser 90 (dark purple), Ser167Phe;

Arg 91 (light blue), Arg168Leu; Pro 226 (green), Pro345Thr; Gly 260 (yellow), Gly382Asp; and Pro 269 (dark pink), Pro391Leu. Image generated from pdb file 1RRQ from the Worldwide Protein Data Bank, based on data from ref. 32. **b**, Alignment of *E. coli* MutY, *B. stearothermophilus* MutY and human MUTYH. Representative mutations that have been observed in individuals with MUTYH-associated polyposis are indicated. The original MUTYH variants, Tyr165Cys and Gly382Asp, are highlighted in red and blue, respectively. Crucial DNA-binding motifs are the helix–hairpin–helix (HhH) motif and the Fe–S cluster loop (FCL) motif^{99,100}. X, stop codon results in truncation at this amino acid.

cell-extract preparations, both human MUTYH variants seemed to be completely inactive⁴³. Taken together, these observations suggest that defects in finding and recognizing damaged DNA might be magnified in a cellular context compared with *in vitro* experiments as a result of the presence of many other DNA-binding enzymes. In addition, the search process is much more demanding in a cellular environment because of the much larger size of chromosomal DNA than the oligonucleotide substrates used *in vitro*. These ideas could explain why MUTYH variants seem to be less active in cellular complementation assays^{37,44}.

Consistent with a global defect in 8-oxoG•A repair, a high proportion of tumours from patients with biallelic mutations in *MUTYH* have been observed to contain G-to-T transversions in the first G of codon 12 of the oncogene *Ki-Ras*⁴⁵. This finding suggests that other genes, in addition to *APC*, might be susceptible to mutation as a result of dysfunctional MUTYH; however, so far, no further compelling connections with other types of cancer have been reported. The observed mutation of *Ki-Ras* in MUTYH-associated polyposis tumours was also interesting, because the same type of inactivating mutation is observed in mice that are deficient in both MUTYH and OGG1 (ref. 46). These double-knockout mice show high susceptibility to tumour formation, mainly lung and ovarian tumours and lymphomas. In addition, these mice also have increased amounts of 8-oxoG in the lungs and small intestine⁴⁷. Surprisingly, mice that are deficient only in MUTYH do not show any atypical properties⁴⁶. However, crossing MUTYH-deficient mice with multiple intestinal neoplasia (*Apc*^{Min/+}) mice, which carry a nonsense mutation in *Apc*, resulted in greater intestinal tumorigenesis than in *Apc*^{Min/+}/*Mutyh*^{+/-} mice⁴⁸. These mouse model studies showed that MUTYH deficiency does result in an increased intrinsic mutation rate; however, this factor alone does not lead to tumour formation in mice. Deficiencies in other proteins, such as APC or OGG1, help to fuel the progression towards cancer. In humans, the extent of oxidative stress and the level of mutations mediated by 8-oxoG, especially in the gastrointestinal tract, might be greater than in mice. It should also be noted that in these mouse models, MUTYH protein is absent, rather than present in a variant form. Studies of *Mutyh*-variant knock-in mice would be interesting, because they would shed light on whether the presence of the variant MUTYH proteins increases mutagenesis and is thereby more deleterious than the complete absence of the protein.

Beyond family N

Since the original discovery of family N, considerable work has established the relationship between mutations in *MUTYH* and colorectal adenomas and carcinomas, and this disorder is now referred to as MUTYH-associated polyposis^{38,49–54}. MUTYH-associated polyposis seems to be an autosomal recessive disorder (requiring mutations in both alleles). Present estimates of the frequency indicate that MUTYH-associated polyposis accounts for ~1% of all colorectal cancer, although this might increase as more patients are tested for mutations in *MUTYH*⁵⁵. This contribution to colorectal cancer is lower than the present estimate of 1–6% for hereditary non-polyposis colorectal cancer (HNPCC)^{56,57}. HNPCC, like MUTYH-associated polyposis, stems from an inherited DNA-repair defect, specifically in genes involved in mismatch repair (MMR)⁵⁷. Of the MMR gene mutations, those in the genes encoding the human MutS homologue, MSH2, and the human MutL homologue, MLH1, predominate⁵⁷. However, HNPCC differs from MUTYH-associated polyposis in being an autosomal dominant disorder and in resulting from a different type of DNA-repair defect. Thus, the evolutionary mechanisms that led to cancer originating from these two DNA-repair defects are not likely to be the same. Although familial adenomatous polyposis and HNPCC are relatively rare, understanding the genetic basis of these disorders has provided important insight into pathways that result in cancer^{58,59}. For example, a feature of the MMR defect is the expansion or contraction of short nucleotide repeats in the DNA of tumours, a process commonly referred to as microsatellite instability^{56,57}. Interestingly, microsatellite instability is also characteristic of many sporadic colorectal cancers⁶⁰. The presence of microsatellite instability, family history, the presence and type

of MMR gene mutation, and clinical manifestations are all important pieces of the puzzle that are used to diagnose HNPCC and to advise and treat patients and family members appropriately⁶¹. Although much is known about HNPCC, many questions still remain about the molecular mechanism that leads from an MMR defect to cancer.

Because MUTYH-associated polyposis was discovered only recently, more information about the clinical and molecular properties of MUTYH-associated polyposis is needed to aid in the diagnosis and treatment of affected patients and family members. Such knowledge might also provide insight into how *MUTYH* mutations contribute more globally to cancer. At present, testing for mutations in *MUTYH* is recommended for patients who have clinical features of familial adenomatous polyposis but either do not have inherited mutations in *APC* or have a family history consistent with recessive inheritance, as is the case for MUTYH-associated polyposis⁵⁰. In classical familial adenomatous polyposis, the sheer number of adenomatous polyps leads to a lifetime risk of eventually developing carcinoma of almost 100% (ref. 36). On the basis of preliminary studies^{53,54,62,63}, a similar lifetime risk for colorectal cancer has been suggested for MUTYH-associated polyposis⁵⁷. The potential increased colorectal cancer risk for heterozygous individuals (who have only one mutated *MUTYH* allele) is a subject of debate, and more clinical testing is required to resolve this issue³⁹.

In addition to the two variants in *MUTYH* that were initially identified (namely Tyr165Cys and Gly382Asp), ~82 germline mutations have been found in the *MUTYH* alleles of patients with colorectal adenomas and carcinomas³⁹ (Fig. 6b). The original variants seem to be the most common mutations found in Caucasians of northern European origin^{38,39}, and different mutations are more prevalent in other ethnic populations³⁹. Although many of these mutations might result in lack of MUTYH, a large number of the 82 mutations (~52) are missense mutations, presenting a diagnostic challenge because their significance is unknown. On the basis of the type of mutation, the location in the sequence of *MUTYH*, and the corresponding position in the structure of the *B. stearothermophilus* MutY-DNA complex, it is possible to make a prediction about the consequences in some cases (Fig. 6a). For example, early truncating mutations (such as Tyr90X, where X denotes that a stop codon results in truncation at this amino acid) in the core of the C-terminal domain or late truncating mutations (such as Glu466X) are likely to be destabilizing and to result in the absence of MUTYH. Some of the missense variants that are conserved in *B. stearothermophilus* MutY might affect DNA binding (for example, Arg168Leu and Ser167Phe) or folding of one of the domains (for example, Trp117Arg). However, the consequences of other conserved missense variants (for example, Pro325Thr and Pro391Leu) are not obvious. Moreover, some missense variants are not conserved in the bacterial MutY enzymes, so it is difficult to make predictions. At this early stage in our understanding of MUTYH-associated polyposis, the possibility that some mutations are non-pathogenic polymorphisms that are found coincidentally in patients with colorectal adenomas and carcinomas cannot be ruled out. It is also possible that some variants are only mildly pathogenic and result in less severe disease symptoms. This idea is supported by *in vitro* results indicating that the Gly382Asp *MUTYH* variant is less catalytically compromised than the Tyr165Cys variant; however, it is unclear whether these findings correlate with clinical data. Reduced A-removal activity has also been reported for four other variants (Arg227Trp, Arg231Leu, Val232Phe and Ala459Asp)^{64–66}. Importantly, mutations in *MUTYH* might also alter *MUTYH* expression levels and stability, protein–protein interactions and post-translational modifications. These and other factors, together with reduced 8-oxoG•A mismatch affinity and A-removal activity, could cooperate to reduce the efficiency of 8-oxoG•A repair in the *APC* gene. Additional clinical and functional data will be particularly important for revealing any associations between colorectal cancer and specific mutations in *MUTYH*.

Why are inherited mutations in *MUTYH* and not other BER enzymes associated with cancer? One possibility is that dysfunction of *MUTYH* is especially problematic because there are no other mechanisms for

repairing 8-oxoG•A mismatches. By contrast, nucleotide excision repair seems to function as a 'back-up' repair mechanism to mediate removal of 8-oxoG in mouse cells lacking OGG1 (refs 67–70). It is also important to note that the clinical feature of having colonic polyps provides a key piece of information that prompts testing for APC or MUTYH mutations, thereby increasing the probability of detecting a relatively rare defective gene. We predict that dysfunction of other BER glycosylases will also be found to modify cancer susceptibility, and these correlations might be uncovered as screening for genetic mutations progresses. Moreover, as information on the molecular basis of cancer becomes available, additional links between BER defects and cancer will be more readily revealed.

Beyond 8-oxoG

Although 8-oxoG has garnered much attention, a plethora of oxidized guanine lesions have been identified¹⁰ (Fig. 2b). Numerous studies indicate that the most common mutations observed in cells in conditions of oxidative stress are G-to-T and G-to-C transversions¹⁰. Whereas G-to-T transversions can be readily explained by the presence of 8-oxoG or FapyG, other oxidized lesions are probably responsible for mediating G-to-C transversions. Of these other lesions, two hydantoin products — spiroiminodihydantoin (Sp) and 5-guanidinohydantoin (Gh) — are particularly interesting and are emerging as important lesions to be dealt with by repair enzymes.

For many years, the product of singlet-oxygen damage to G was thought to be 4-OH-8-oxoG^{71,72}, and the presence of this product was used routinely as an indicator of singlet-oxygen damage⁷³. However, when Burrows and co-workers determined the structure of Sp, as a major product arising from further oxidation of 8-oxoG-containing nucleosides and oligonucleotides, suspicion was raised about the structure of 4-OH-8-oxoG, because the molecular weights are the same⁷⁴. It was later confirmed that the structure of 4-OH-8-oxoG had been misassigned and that 4-OH-8-oxoG was Sp^{75,76}. When the identity of Sp was unmasked, it was observed to be generated by oxidation of G or 8-oxoG by a large number of oxidants, including singlet oxygen, high-valent metal ions, peroxy nitrite and ionizing radiation^{10,77}. Interestingly, the studies of oxidation of 8-oxoG revealed that, depending on the conditions, Gh is formed in addition to Sp⁷⁸. When the substrate is a nucleoside or single-stranded DNA, the main product is Sp, whereas Gh predominates in double-stranded DNA⁷⁷.

Hydantoin lesions have been demonstrated to be highly mutagenic both *in vitro* and *in vivo*. Single-nucleotide insertion and primer extension experiments using an *E. coli* Klenow fragment of DNA polymerase without exonuclease activity indicate that dAMP and dGMP are inserted opposite these oxidized lesions^{79,80}. In *E. coli*-based mutagenesis assays with single-stranded lesion-containing viral DNA, Sp was found to be a strong block to replication, whereas Gh was more readily bypassed^{81,82}. Moreover, both lesions are 100% mutagenic, mediating both G-to-C and G-to-T transversions, whereas 8-oxoG is only mildly mutagenic (3%) in these assays^{81,82}.

Important features that influence the mutagenic potential of a given lesion are the efficiency of its repair and the sensitivity of the repair enzymes to the correct base-pairing context. The relatively low mutation frequency of 8-oxoG in both *E. coli* and mammalian cells results from the efficient repair of 8-oxoG. *In vitro* assays have shown that Gh and Sp are substrates for *E. coli* Fpg (MutM)⁸³. Efficient repair in the correct base-pairing context would be expected to mitigate the mutagenic potential of lesions; however, removal in the wrong context might increase mutagenesis. This is shown by studies of Gh, which was found to be excised by Fpg from Gh•G and Gh•C base pairs with similar efficiencies. The Gh•G base pair is particularly interesting because it might mediate the G-to-C transversions observed *in vivo*. Removal of Gh or Sp opposite A by Fpg was less than that opposite C; however, the extent of removal of these oxidized lesions is greater than removal of 8-oxoG opposite A⁸³. The activity in base-pairing contexts involving A might also be more problematic, because MutY was found to be unable to remove A from Gh•A or Sp•A base pairs⁸³.

Surprisingly, Gh and Sp are not substrates for OGG1, although they are removed by the *Saccharomyces cerevisiae* homologues Ogg1 and Ntg1 (ref. 84). Another bacterial glycosylase, endonuclease VIII (Nei), can also remove Gh and Sp⁸⁵. Interestingly, this enzyme usually targets oxidized pyrimidines, such as thymine glycol and 5-hydroxycytosine⁸⁶. In fact, Gh and Sp, like 8-oxoG, have a thymine-like Watson–Crick hydrogen-bonding face (Fig. 2b). On the basis of the idea that repair might thwart the ability to detect lesions that are present at low levels in cells, Sugden and co-workers used various repair-deficient *E. coli* strains to identify the products formed by chromate oxidation⁸⁷; Sp was detected by mass spectrometric techniques in chromate-treated cells deficient in the BER glycosylase Nei.

Recently, mammalian orthologues of *E. coli* Fpg and Nei were identified and designated as the Nei-like (NEIL) family of enzymes (which consists of NEIL1, NEIL2 and NEIL3)^{88–91}. The overall structure of NEIL1 is similar to that of *E. coli* Nei and Fpg, although it contains an unusual 'zinless'-finger motif, which is required for its glycosylase activity⁹². The substrate specificity of NEIL1 is more similar to that of Nei than that of Fpg⁸⁸. Reported substrates include 5-hydroxyuracil, Fapy nucleotides and thymine glycol, whereas the activity towards 8-oxoG is minimal. In contrast to OGG1, NEIL1 also operates on single-stranded DNA and 'bubble DNA' (that is, single-stranded DNA flanked by duplex regions), thus prompting the suggestion that it is involved in replication and/or transcription-coupled repair⁹³. Qualitative reports also showed that Gh and Sp lesions are removed by the mouse counterparts of Nei, NEIL1 and NEIL2 (ref. 94). Interestingly, quantitative examination of the substrate specificity of human NEIL1 revealed Gh and Sp lesions as the best substrates identified so far (N. Krishnamurthy and S.S.D., unpublished observations). Notably, mouse and human NEIL1 have been found to remove Gh and Sp from all base-pairing contexts that would be mutagenic, except when paired with C. The detection of Sp lesions in cells, together with the high mutagenic potential of both Gh and Sp and the potentially muddled processes of repair, suggests that these lesions will begin to capture more and more of the attention that is presently directed toward 8-oxoG.

The importance of human NEIL1 in response to oxidative stress has been illustrated by the observation of increased amounts of *NEIL1* mRNA in response to ROS⁹⁵. Moreover, knockdown of NEIL1 using RNA interference produces cells that are extremely sensitive to ionizing radiation⁹⁶. In addition, an intriguing report correlated inactivating mutations in *NEIL1* with human gastric cancer⁹⁷. These varied biological consequences are not surprising for an enzyme involved in the repair of oxidative damage; however, some peculiarities in the biological role of human NEIL1 were revealed in studies of the recently generated *Neil1*-knockout mouse⁹⁸. In the absence of oxidative stress, both *Neil1*-knockout mice (*Neil1*^{-/-}) and mice that are heterozygous for *Neil1* (*Neil1*^{+/−}) show symptoms consistent with metabolic syndrome in humans⁹⁸. These symptoms include severe obesity, dyslipidaemia and fatty liver disease. It seems somewhat odd for a defect in a BER glycosylase to be correlated with a metabolic disorder, but these types of disorder have been linked to oxidative stress. The authors of that study, Vartanian and co-workers, suggest that disruption of energy homeostasis consistent with metabolic disorders might result from the extensive mitochondrial DNA damage observed in NEIL1-deficient mice⁹⁸. Alternatively, damage to nuclear DNA in specific cell types, such as liver and pancreatic cells, might be extensive in the absence of NEIL1, leading to symptoms of metabolic syndrome. Clearly, these interesting results warrant further study, and they illustrate the important role of BER in the overall well-being of organisms.

Concluding remarks

Since the discovery of BER more than three decades ago, the field has clearly taken its place in the mainstream of DNA repair. A plethora of structural information has provided amazing insight into the processes of DNA-damage recognition and excision by these enzymes. This structural work will probably be complemented in the future by new biophysical and biochemical approaches allowing direct visualization of

the search processes by a variety of DNA glycosylases. This could allow additional intermediates to be observed and provide a more generalized model for damage recognition by these enzymes.

Uncovering the molecular details of events involved in the processing of damaged DNA bases by DNA glycosylases will be important to understand further how the aberrant function of these enzymes causes disease. Here, this relationship was illustrated by discussing MUTYH variants, for which the biochemical and structural data are consistent with a reduced efficiency to recognize and repair 8-oxoG•A mismatches. This provides a molecular basis for the observed G-to-T transversion mutations in *APC*, leading to MUTYH-associated polyposis. Understanding of MUTYH-associated polyposis is still at an early stage, however, and many unanswered questions about the role of the variant enzymes in initiating the process of carcinogenesis remain. For example, does severity of the disease correlate with the type of *MUTYH* mutation? Why are the colon and the *APC* gene particularly sensitive to a BER defect? The intriguing metabolic disorder phenotype of the *Neil1*-knockout mouse also suggests that interesting relationships between BER and human disease will continue to be discovered. Our sophisticated understanding of the intricate steps used to recognize and excise aberrant DNA bases could now be exploited to selectively target steps in the 'search and rescue' process mediated by DNA glycosylases. For example, small-molecule modulators of glycosylase activity could be new tools for chemical biology studies and might lead to new therapeutic approaches. ■

1. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
2. Friedberg, E. C. DNA damage and repair. *Nature* **421**, 436–440 (2003).
3. Pfeifer, G. P. *et al.* Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**, 7435–7451 (2002).
4. Friedberg, E. C. Inroads into base excision repair II. The discovery of the DNA glycosylases. *DNA Repair (Amst.)* **3**, 1531–1536 (2004).
5. David, S. S. & Williams, S. D. Chemistry of glycosylases and endonucleases involved in base-excision repair. *Chem. Rev.* **98**, 1221–1261 (1998).
6. Fromme, J. C. & Verdine, G. L. Base excision repair. *Adv. Protein Chem.* **69**, 1–41 (2004).
7. Barnes, D. E. & Lindahl, T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu. Rev. Genet.* **35**, 445–476 (2004).
8. Sung, J.-S. & Demple, B. Roles of base excision repair subpathways in correcting oxidized abasic sites in DNA. *FEBS J.* **273**, 1620–1629 (2006).
9. Klaunig, J. E. & Kamendulis, L. M. The role of oxidative stress in carcinogenesis. *Annu. Rev. Pharmacol. Toxicol.* **44**, 239–267 (2004).
10. Neely, W. L. & Essigmann, J. M. Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chem. Res. Toxicol.* **19**, 491–505 (2006).
11. Burrows, C. M. & Muller, J. Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* **98**, 1109–1152 (1998).
12. Shibusaki, S., Takeshita, M. & Grollman, A. P. Insertion of specific bases during DNA synthesis past the oxidation damaged base 8-oxodG. *Nature* **349**, 431–434 (1991).
13. Hsu, G. W., Ober, M., Carell, T. & Beese, L. S. Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* **431**, 217–221 (2004).
14. Michaels, M. L. & Miller, J. H. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.* **174**, 6321–6325 (1992).
15. Krah, J. M., Beard, W. A., Miller, H., Grollman, A. P. & Wilson, S. H. Structure of DNA polymerase β with the mutagenic DNA lesion 8-oxodeoxyguanine reveals structural insights into its coding potential. *Structure* **11**, 121–127 (2003).
16. Gedik, C. M. & Collins, A. Establishing the background level of base oxidation in human lymphocyte DNA: results on an interlaboratory validation study. *FASEB J.* **19**, 82–84 (2005).
17. Parikh, S. S., Putnam, C. D. & Tainer, J. A. Lessons learned from structural results on uracil-DNA glycosylase. *Mutat. Res.* **460**, 183–199 (2000).
18. Stivers, J. T. Site-specific DNA damage recognition by enzyme-induced base flipping. *Prog. Nucleic Acid Res. Mol. Biol.* **77**, 37–65 (2004).
19. Fromme, J. C., Banerjee, A. & Verdine, G. L. DNA glycosylase recognition and catalysis. *Curr. Opin. Struct. Biol.* **14**, 43–49 (2004).
20. Huffman, J. L., Sundheim, O. & Tainer, J. A. DNA base damage recognition and removal: new twists and grooves. *Mutat. Res.* **577**, 55–76 (2005).
21. Hitomi, K., Iwai, S. & Tainer, J. A. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal and repair. *DNA Repair (Amst.)* **6**, 410–428 (2007).
22. Bruner, S. D., Norman, D. P. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866 (2000).
23. Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612–618 (2005).
24. Radom, C. T., Banerjee, A. & Verdine, G. L. Structural characterization of human 8-oxoguanine DNA glycosylase variants bearing active site mutations. *J. Biol. Chem.* **282**, 9182–9194 (2007).
25. Banerjee, A. & Verdine, G. L. A nucleobase lesion remodels the interaction of its normal neighbor in a DNA glycosylase complex. *Proc. Natl. Acad. Sci. USA* **103**, 15020–15025 (2006).
26. Banerjee, A., Santos, W. L. & Verdine, G. L. Structure of a DNA glycosylase searching for DNA lesions. *Science* **311**, 1153–1157 (2006).
27. Fromme, J. C. & Verdine, G. L. DNA lesion recognition by the bacterial repair enzyme MutM. *J. Biol. Chem.* **278**, 51543–51548 (2003).
28. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. USA* **103**, 5752–5757 (2006).
29. Jiang, Y. L. *et al.* Recognition of an unnatural difluorophenyl nucleotide by uracil DNA glycosylase. *Biochemistry* **43**, 15429–15438 (2004).
30. Noll, D. M., Gogos, A., Granek, J. A. & Clarke, N. D. The C-terminal domain of the adenine-DNA glycosylase MutY confers specificity of 8-oxoguanine–adenine mispairs and may have evolved from MutT, an 8-oxo-dGTPase. *Biochemistry* **38**, 6374–6379 (1999).
31. Chmiel, N. H., Golinelli, M.-P., Francis, A. W. & David, S. S. Efficient recognition of substrates and substrate analogs by the adenine glycosylase MutY requires the C-terminal domain. *Nucleic Acids Res.* **29**, 553–564 (2001).
32. Fromme, J. C., Banerjee, A., Huang, S. J. & Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **427**, 652–656 (2004).
33. Wiederhold, C. J., Delaney, M. O., Pope, M. A., David, S. S. & Greenberg, M. M. Repair of DNA containing FapydG and its C-nucleoside analogue by formamidopyrimidine DNA glycosylase and MutY. *Biochemistry* **42**, 9755–9760 (2003).
34. Bernards, A. S., Miller, J. K., Bao, K. K. & Wong, I. Flipping duplex DNA inside out: a double base-flipping reaction mechanism by *Escherichia coli* MutY adenine glycosylase. *J. Biol. Chem.* **277**, 20960–20964 (2002).
35. Al-Tassan, N. *et al.* Inherited variants of MYH associated with somatic G:C to T:A mutations in colorectal tumors. *Nature Genet.* **30**, 227–232 (2002).
36. Fearnhead, N. S., Britton, M. P. & Bodmer, W. F. The ABC of APC. *Hum. Mol. Genet.* **10**, 721–733 (2001).
37. Chmiel, N. H., Livingston, A. L. & David, S. S. Insight into the functional consequences of inherited variants of the hMYH adenine glycosylase associated with colorectal cancer: complementation assays with hMYH variants and pre-steady-state kinetics of the corresponding mutated *E. coli* enzymes. *J. Mol. Biol.* **327**, 431–443 (2003).
38. Sampson, J. R., Jones, S., Dolwani, S. & Cheadle, J. P. MutYH (MYH) and colorectal cancer. *Biochem. Soc. Trans.* **33**, 679–683 (2005).
39. Cheadle, J. P. & Sampson, J. R. MUTYH-associated polyposis — from defect in base excision repair to clinical genetic testing. *DNA Repair (Amst.)* **6**, 274–279 (2007).
40. Livingston, A. L., Kundu, S., Henderson-Pozzi, M., Anderson, D. W. & David, S. S. Insight into the roles of tyrosine 82 and glycine 253 in the *Escherichia coli* adenine glycosylase MutY. *Biochemistry* **44**, 14179–14190 (2005).
41. Pope, M. A., Chmiel, N. H. & David, S. S. Insight into the functional consequences of hMYH variants associated with colorectal cancer: distinct differences in the adenine glycosylase activity and the response to AP endonuclease of Y150C and G365D murine MYH. *DNA Repair (Amst.)* **4**, 315–325 (2005).
42. Tominaga, Y. *et al.* MUTYH prevents OGG1 or APEX1 from inappropriately processing its substrate or reaction product with its C-terminal domain. *Nucleic Acids Res.* **32**, 3198–3211 (2004).
43. Wooden, S. H., Bassett, H. M., Wood, T. G. & McCullough, A. K. Identification of critical residues required for the mutation avoidance function of human MutY (hMYH) and implications in colorectal cancer. *Cancer Lett.* **205**, 89–95 (2004).
44. Hirano, S. *et al.* Mutator phenotype of MutYH-null mouse embryonic stem cells. *J. Biol. Chem.* **278**, 38121–38124 (2003).
45. Lipton, L. *et al.* Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. *Cancer Res.* **63**, 7595–7599 (2003).
46. Xie, Y. L. *et al.* Deficiencies in mouse Myh and Ogg1 results in tumor predisposition and G to T mutations in codon 12 of the K-ras oncogene in lung tumors. *Cancer Res.* **64**, 3096–3102 (2004).
47. Russo, M. T. *et al.* Accumulation of the oxidative base lesion 8-hydroxyguanine in DNA of tumor-prone mice defective in both the Myh and Ogg1 DNA glycosylase. *Cancer Res.* **64**, 4411–4414 (2004).
48. Sieber, O. M. *et al.* Myh deficiency enhances intestinal tumorigenesis in multiple intestinal neoplasia (*Apc*^{Min/+}) mice. *Cancer Res.* **64**, 8876–8881 (2004).
49. Sampson, J. R. *et al.* MYH polyposis: a new autosomal recessive form of familial adenomatous polyposis demanding reappraisal of genetic risk and family management. *Lancet* **362**, 39–41 (2003).
50. Chow, E., Thirlwell, C., Macrae, F. & Lipton, L. Colorectal cancer and inherited mutations in base-excision repair. *Lancet Oncol.* **5**, 600–606 (2004).
51. Lipton, L. & Tomlinson, I. The multiple colorectal adenoma phenotype and MYH, a excision repair gene. *Clin. Gastroenterol. Hepatol.* **2**, 633–638 (2004).
52. Tenesa, A. *et al.* Association of MutYH and colorectal cancer. *Br. J. Cancer* **95**, 239–242 (2006).
53. Lipton, L. & Tomlinson, I. The genetics of FAP and FAP-like syndromes. *Fam. Cancer* **5**, 221–226 (2006).
54. Farrington, S. M. *et al.* Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am. J. Hum. Genet.* **77**, 112–119 (2005).
55. Fleischmann, C. *et al.* Comprehensive analysis of the contribution of germline MYH variation of early-onset colorectal cancer. *Int. J. Cancer* **109**, 554–558 (2004).
56. Strate, L. L. & Syngal, S. Hereditary colorectal cancer syndromes. *Cancer Causes Control* **16**, 201–213 (2005).
57. Jo, W. S. & Chung, D. C. Genetics of hereditary colorectal cancer. *Semin. Oncol.* **32**, 11–23 (2005).
58. Bodmer, W. F. Cancer genetics: colorectal cancer as a model. *J. Hum. Genet.* **51**, 391–396 (2006).
59. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
60. Soreide, K., Janssen, E. A. M., Soiland, H., Korner, H. & Baak, J. P. Microsatellite instability in colorectal cancer. *Br. J. Surg.* **93**, 395–406 (2006).
61. Lindor, N. M. *et al.* Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome. *JAMA* **296**, 1507–1517 (2006).

62. Venesio, T. *et al.* High frequency of MYH gene mutations in a subset of patients with familial adenomatous polyposis. *Gastroenterology* **126**, 1681–1685 (2004).

63. Leite, J. S. *et al.* Is prophylactic colectomy indicated in patients with MYH-associated polyposis? *Colorectal Dis.* **7**, 327–331 (2005).

64. Bai, H. *et al.* Functional characterization of two human MutY homolog (hMYH) missense mutations (R227W and V232F) that lie within the putative hMSH6 binding domain and are associated with hMYH polyposis. *Nucleic Acids Res.* **33**, 597–604 (2005).

65. Bai, H. *et al.* Functional characterization of human MutY homolog (hMYH) missense mutation (R231L) that is linked with hMYH-associated polyposis. *Cancer Lett.* **250**, 74–81 (2007).

66. Alhopuro, P. *et al.* A novel functionally deficient MYH variant in individuals with colorectal adenomatous polyposis. *Hum. Mutat.* **26**, 393 (2005).

67. Klunglund, A. *et al.* Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proc. Natl Acad. Sci. USA* **96**, 13300–13305 (1999).

68. Osterod, M. *et al.* A global DNA repair mechanism involving Cockayne syndrome B (CSB) gene product can prevent the *in vivo* accumulation of endogenous oxidative DNA base damage. *Oncogene* **21**, 8232–8239 (2002).

69. Osterod, M. *et al.* Age-related and tissue-specific accumulation of oxidative DNA base damage in 7,8-dihydro-8-oxoguanine-DNA glycosylase (*Ogg1*) deficient mice. *Carcinogenesis* **22**, 1459–1463 (2001).

70. Sunesen, M., Stevnsner, T., Brosh, R. M., Dianov, G. L. & Bohr, V. A. Global genome repair of 8-oxoG in hamster cells requires a functional CSB gene product. *Oncogene* **21**, 3571–3578 (2002).

71. Cadet, J., Decarroz, C., Wang, S. Y. & Midden, W. R. Mechanisms and products of photosensitized degradation of nucleic acids and related model compounds. *Isr. J. Chem.* **1983**, 420–429 (1983).

72. Ravanat, J. L. & Cadet, J. Reaction of singlet oxygen with 2'-deoxyguanosine and DNA. Isolation and characterization of the main oxidation products. *Chem. Res. Toxicol.* **8**, 379–388 (1995).

73. Ravanat, J. L., Berger, M., Bernard, F., Langlois, R. & Ouellet, R. Phthalocyanine and naphthalocyanine photosensitized oxidation of 2'-deoxyguanosine: distinct type I and type II products. *Photochem. Photobiol.* **55**, 809–814 (1992).

74. Luo, W., Muller, J. G., Rachlin, E. M. & Burrows, C. J. Characterization of spiroiminodihydantoin as a product of one-electron oxidation of 8-oxo-7,8-dihydroguanosine. *Org. Lett.* **2**, 613–616 (2000).

75. Niles, J. C., Wishnok, J. S. & Tannenbaum, S. R. Spiroiminodihydantoin is the major product of the 8-oxo-7,8-dihydroguanosine reaction with peroxynitrite in the presence of thiols and guanosine photooxidation by methylene blue. *Org. Lett.* **3**, 963–966 (2001).

76. Adam, W. *et al.* Spiroiminodihydantoin is a major product in the photooxidation of 2'-deoxyguanosine by the triplet states and oxyl radicals generated from hydroxyacetophenone photolysis and dioxetane thermolysis. *Org. Lett.* **4**, 537–540 (2002).

77. Burrows, C. J. *et al.* Structure and potential mutagenicity of new hydantoin products from guanosine and 8-oxo-7,8-dihydroguanosine oxidation by transition metals. *Environ. Health Perspect.* **110**, 713–717 (2002).

78. Luo, W., Muller, J. G., Rachlin, E. M. & Burrows, C. J. Characterization of hydantoin products from one-electron oxidation of 8-oxo-7,8-dihydroguanosine in a nucleoside model. *Chem. Res. Toxicol.* **14**, 927–938 (2001).

79. Korniyushyna, O., Berges, A. M., Muller, J. G. & Burrows, C. J. *In vitro* nucleotide misinsertion opposite the oxidized guanosine lesions spiroiminodihydantoin and guanidinohydantoin and DNA synthesis past the lesions using *Escherichia coli* DNA polymerase I (Klenow fragment). *Biochemistry* **41**, 15304–15314 (2002).

80. Korniyushyna, O. & Burrows, C. J. Effect of the oxidized lesions spiroiminodihydantoin and guanidinohydantoin on proofreading by *Escherichia coli* DNA polymerase I (Klenow fragment) in different sequence contexts. *Biochemistry* **42**, 13008–13018 (2003).

81. Henderson, P. T. *et al.* The hydantoin lesions from oxidation of 7,8-dihydro-8-oxoguanine are potent sources of replication errors *in vivo*. *Biochemistry* **42**, 9257–9262 (2003).

82. Delaney, S., Neeley, W. L., Delaney, J. C. & Essigmann, J. M. The substrate specificity of MutY for hyperoxidized guanine lesions *in vivo*. *Biochemistry* **46**, 1448–1455 (2007).

83. Leipold, M. D., Muller, J. G., Burrows, C. J. & David, S. S. Removal of hydantoin products of 8-oxoguanine oxidation by the *Escherichia coli* DNA repair enzyme, Fpg. *Biochemistry* **39**, 14984–14992 (2000).

84. Leipold, M. D., Workman, H., Muller, J. G., Burrows, C. J. & David, S. S. Recognition and removal of oxidized guanines in duplex DNA by the base excision repair enzymes hOGG1, yOGG1 and yOGG2. *Biochemistry* **42**, 11373–11381 (2003).

85. Hazra, T. K. *et al.* Repair of hydantoins, one electron oxidation product of 8-oxoguanine, by DNA glycosylases of *Escherichia coli*. *Nucleic Acids Res.* **29**, 1967–1974 (2001).

86. Wallace, S. S., Bandaru, V., Kathe, S. D. & Bond, J. P. The enigma of endonuclease VIII. *DNA Repair (Amst.)* **2**, 441–453 (2003).

87. Hailer, M. K., Slade, P. G., Martin, B. D. & Sugden, K. D. Nei-deficient *Escherichia coli* are sensitive to chromate and accumulate the oxidized guanine lesion spiroiminodihydantoin. *Chem. Res. Toxicol.* **18**, 1378–1383 (2005).

88. Bandaru, V., Sunkara, S., Wallace, S. S. & Bond, J. P. A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair (Amst.)* **1**, 517–529 (2002).

89. Hazra, T. K. *et al.* Identification and characterization of a human DNA glycosylase for repair of modified oxidatively damaged DNA. *Proc. Natl Acad. Sci. USA* **99**, 3523–3528 (2002).

90. Hazra, T. K. *et al.* Identification of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J. Biol. Chem.* **277**, 30417–30420 (2002).

91. Morland, I. *et al.* Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res.* **30**, 4926–4936 (2002).

92. Double, S., Bandaru, V., Bond, J. P. & Wallace, S. S. The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zinc-finger motif required for glycosylase activity. *Proc. Natl Acad. Sci. USA* **101**, 10284–10289 (2004).

93. Dou, H., Mitra, S. & Hazra, T. K. Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J. Biol. Chem.* **278**, 49679–49684 (2003).

94. Hailer, K. M., Slade, P. G., Martin, B. D., Rosenquist, T. A. & Sugden, K. D. Recognition of the oxidized lesions spiroiminodihydantoin and guanidinohydantoin in DNA by the base excision repair glycosylases NEIL1 and NEIL2. *DNA Repair (Amst.)* **4**, 41–50 (2005).

95. Das, A., Hazra, T. K., Boldogh, I., Mitra, S. & Bhakat, K. K. Induction of the human oxidized base-specific DNA glycosylase NEIL1 by reactive oxygen species. *J. Biol. Chem.* **280**, 35272–35280 (2005).

96. Rosenquist, T. A. *et al.* The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. *DNA Repair (Amst.)* **2**, 581–591 (2003).

97. Shimura, K. *et al.* Inactivating mutations of the human base excision repair gene NEIL1 in gastric cancer. *Carcinogenesis* **25**, 2311–2317 (2004).

98. Vartanian, V. *et al.* The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase. *Proc. Natl Acad. Sci. USA* **103**, 1864–1869 (2006).

99. Guan, Y. *et al.* MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nature Struct. Biol.* **5**, 1058–1064 (1998).

100. Lukianova, O. L. & David, S. S. A role for iron-sulfur clusters in DNA repair. *Curr. Opin. Chem. Biol.* **9**, 145–151 (2005).

Acknowledgements We thank members of the David laboratory for reading the manuscript. We also apologize to all scientists whose original studies and reviews were not included because of space limitations. Research in the laboratory of S.S.D. is funded by the National Institutes of Health, and V.L.O. has been supported by pre-doctoral fellowships from the National Institutes of Health.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to S.S.D. (david@chem.ucdavis.edu).

Chromatin dynamics and the preservation of genetic information

Jessica A. Downs¹, Michel C. Nussenzweig² & André Nussenzweig³

The integrity of the genome is frequently challenged by double-strand breaks in the DNA. Defects in the cellular response to double-strand breaks are a major cause of cancer and other age-related pathologies; therefore, much effort has been directed at understanding the enzymatic mechanisms involved in recognizing, signalling and repairing double-strand breaks. Recent work indicates that chromatin — the fibres into which DNA is packaged with a proteinaceous structural polymer — has an important role in initiating, propagating and terminating this cellular response to DNA damage.

DNA repair is important for preserving the information in our genomes, so this area has been studied intensively for decades. Biochemical studies of DNA repair are often carried out using naked DNA substrates, but this is not physiologically relevant. In a cell, naked DNA is vulnerable to nuclelease digestion and other insults, and without some type of organization, it would occupy more volume than necessary. Consequently, eukaryotic cells compact DNA into chromatin. The repair machinery must therefore deal with breaks in the DNA being relatively inaccessible. In this review, we focus on recent progress in understanding the effects of chromatin on DNA repair and how this potential barrier is relieved to enable repair of double-strand breaks (DSBs).

Organization of the genetic material into chromatin

The basic unit of chromatin is the nucleosome core particle, which consists of ~146 base pairs of DNA wrapped in two left-handed superhelical turns around an octamer of histone proteins (Fig. 1). The histone octamer contains two copies of each of the four conserved core histones: H2A, H2B, H3 and H4. Each protein has both a histone fold domain, which mediates the histone–histone and histone–DNA interactions that are crucial for the assembly of the nucleosome core particle, and a flexible amino-terminal tail domain, which protrudes from the nucleosome core particle. H2A and H2B also have an important carboxy-terminal tail domain (Fig. 1).

Eukaryotes have several histone variants, which, as a result of their altered amino-acid composition, can affect both the structure of individual nucleosomes and the ability of nucleosomes to form higher-order chromatin structure. Moreover, binding of proteins such as linker histones and high-mobility group (HMG) proteins affects the architecture and the degree of chromatin compaction.

Chromatin structure can also be altered by enzymes that remodel chromatin or covalently modify histones. The term chromatin remodelling encompasses a wide variety of changes in chromatin structure but can be defined as a discernible change in histone–DNA contacts. Such changes in contacts can result from the repositioning (sliding) of nucleosomes on DNA, the removal of part or all of the histone octamer from DNA, an induced change in the accessibility of the DNA in chromatin to proteins such as transcription factors or nucleases, and the exchange of histone variants for core histones.

An important enzymatic mechanism for altering chromatin structure involves the covalent modification of histones. This is carried

out by a wide variety of enzymes, including histone acetyltransferases (HATs), histone methyltransferases, protein kinases, and ubiquitin- and SUMO-protein ligases. The combination of modifications ('marks') produced by these enzymes has been proposed to constitute a code that regulates downstream processes such as transcription, DNA repair and apoptosis^{1,2}.

Covalent modification of histone tails can itself influence the state of chromatin compaction³; however, the main mechanism by which these modifications influence chromatin structure is by controlling the binding of non-histone effector proteins to chromatin, some of which (for example, remodellers) have the capacity to alter histone–DNA contacts. Such effector proteins have modules that recognize specific histone modifications. For example, bromodomains bind acetylated lysine residues, whereas chromodomains, tudor domains and plant homeodomain (PHD)-finger domains bind distinct methylated residues.

Higher-order chromatin packaging is a barrier to the detection and repair of DNA damage. However, a study carried out two decades ago indicated that DSBs cause chromatin reorganization, as shown by altered access of nucleases to damaged DNA⁴. More recently, it was shown that DSBs induce a local decrease in the density of the chromatin fibre, in addition to altering the position of nucleosomes and causing the eviction of the histone octamer^{5–7}. These remodelling events can make the lesion more accessible to 'damage sensor' proteins by exposing features of the nucleosome, such as constitutive histone modifications, that are normally concealed in the unperturbed cell⁸.

In addition to remodelling chromatin and unmasking histone modifications buried in the histone fold domains, DSBs also elicit post-translational modifications on the protruding histone tails (Table 1).

The earliest, and arguably the most robust, of these modifications is phosphorylation of the histone H2A variant H2AX (also known as H2AFX) on its extended C-terminal tail. Within seconds, phosphorylated H2AX (known as γ-H2AX) spreads over a region spanning thousands to millions of bases surrounding a DSB⁹. H2AX is not present in lower eukaryotes. However, the motif that is phosphorylated in response to DNA damage is present in the C-terminal tails of other H2A-family members. For example, in *Saccharomyces cerevisiae*, it is present in the core histone H2A, and in *Drosophila melanogaster*, it is present in the H2AZ-family member H2AV (also known as HIS2AV).

¹MRC Genome Damage and Stability Centre, University of Sussex, Brighton BN1 9RQ, UK. ²Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA. ³Experimental Immunology Branch, National Cancer Institute, National Institutes of Health, 10 Center Drive, 4B04, Bethesda, Maryland 20892, USA.

The cellular response to DSBs

There are two main pathways for repairing DSBs: non-homologous end joining (NHEJ) (Fig. 2A) and homologous recombination (Fig. 2B). NHEJ involves the direct religation of broken DNA with minimal processing of sequences near the ends; significant sequence homology is not required. By contrast, the homologous-recombination repair pathway copies matched base pairs from a complementary template found on the homologous chromosome or sister chromatid. In mammalian cells, NHEJ has a dominant role in repairing DSBs in the G1 phase of the cell cycle, whereas homologous recombination is used predominantly in S phase, concomitantly with the appearance of the sister chromatid template.

DSBs are usually repaired rapidly, but when a DSB persists, a DNA-damage checkpoint is triggered^{10,11}. The checkpoint arrests the cell cycle, providing a prolonged period during which further attempts can be made to fix the damage. When the lesion is repaired, the arrest signal is extinguished, and cells re-enter the cell cycle. If the lesion is not repairable, then the cell is programmed to die or undergoes a permanent arrest, thereby preventing propagation of the lesion and potential mutagenesis.

The cellular response to DSBs is initiated and enforced by the activity of phosphatidylinositol-3-OH-kinase-like kinases (PIKKs), principally ATM (ataxia-telangiectasia mutated; Tel1 in *S. cerevisiae*), ATR (ATM and Rad3-related; Mec1 in *S. cerevisiae*) and DNA-dependent protein kinase (DNA-PK; also known as PRKDC). PIKKs are loaded onto the lesion through association with adaptor proteins that have a high affinity for either DSBs that are unprocessed (or minimally processed) or DSBs that have been resected on one strand to form single-stranded DNA (ssDNA) (Fig. 2C).

Oligomeric ATM is loaded onto blunt or minimally processed DSBs through an interaction with the DSB sensor complex MRN, which consists of MRE11 (meiotic recombination 11 homologue), RAD50 and NBS1 (Nijmegen breakage syndrome 1). Two models have been proposed to explain how ATM is activated and recruited to DSBs. One possibility is that, after chromatin unfolding, ATM is converted to an active monomeric form by autophosphorylation and, subsequently, becomes competent for DNA binding¹². In this model, before it is recruited to the DSB, ATM is activated at a distance from the DSB by undefined perturbations in chromatin structure. Alternatively, both recruitment and activation of ATM might depend on the prior loading

of MRN onto the DSB (Fig. 2C). This second hypothesis posits that the enzymatic activity of MRE11 and RAD50 converts ATM to a catalytically active configuration at the site of the lesion^{13,14}. In this case, the autophosphorylation of ATM, similar to its transphosphorylation of downstream targets, would occur after activation at sites of DSBs^{13,15,16}. Local activation of ATM at the site of DNA damage would be analogous to the mechanism responsible for autophosphorylation of the DNA-PK catalytic subunit (DNA-PKcs), a process that results from the recruitment of DNA-PKcs to DSBs by Ku¹⁷ (Fig. 2C).

The molecular species that is recognized by ATR seems to be ssDNA flanked by double-stranded DNA¹⁸. The recruitment of ATR to this form of damaged DNA depends on the ssDNA-binding protein replication protein A (RPA) and on the interaction of ATR with the adaptor ATR-interacting protein (ATRIP; Ddc2 in *S. cerevisiae*, and Rad26 in *Schizosaccharomyces pombe*)¹⁸. ATR is activated on association with DNA, by the mediator topoisomerase-IIβ-binding protein 1 (TOPBP1), which might induce a conformational change in DNA-bound ATR¹⁹ (Fig. 2C).

In addition to PIKK-target phosphorylation events at DSBs, the other prominent feature of the DNA-damage response is the large-scale accumulation of proteins near the DSB site²⁰. Proteins involved in DNA repair, cell-cycle checkpoints and chromatin remodelling (described in detail later) form aggregations that are cytologically detectable as nuclear foci²¹. The dynamic assembly of proteins into foci requires that H2AX is phosphorylated first (a process that is itself mediated by the PIKKs) in a large region surrounding the lesion²⁰. Nevertheless, loss of H2AX does not abrogate DNA-damage checkpoints or repair, probably because the initial recognition of the lesion and the PIKK-mediated phosphorylation of effector proteins do not strictly depend on their accumulation distal to the DSB site²². However, deletion of the gene encoding H2AX impairs the joining of programmed DNA lesions (which are separated by up to 200 kilobases of intervening DNA) during immunoglobulin class-switch recombination, suggesting that chromatin modifications at a distance are required for synapsis^{23,24}.

Chromatin-remodelling activities in DSB repair

Chromatin-remodelling activities seem to facilitate the access of repair and checkpoint proteins to the DNA lesion. In addition, the unfolding of nearby chromatin (rather than the DSB itself) might be the trigger

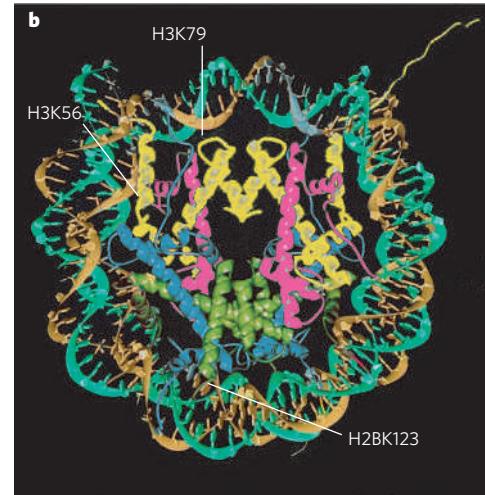
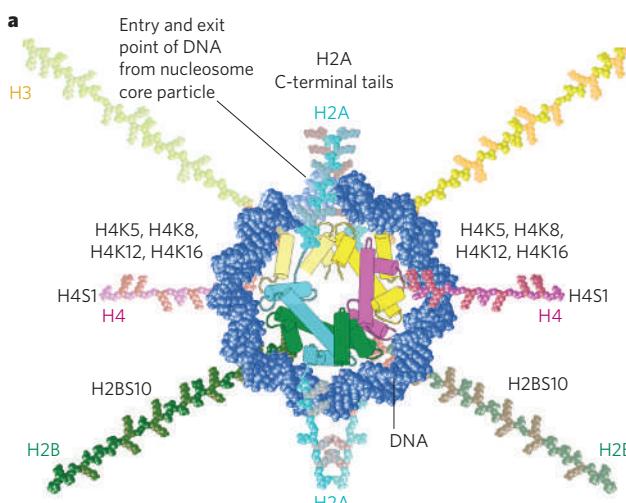


Figure 1 | Structure of the nucleosome core particle and modifications of histones. **a**, Model of the nucleosome core particle. Each nucleosome core particle consists of double-stranded DNA (dark blue) wrapped around two molecules of each of the four core histone proteins: H2A (light blue), H2B (green), H3 (yellow) and H4 (pink). Note that the C-terminal tails of the two H2A proteins protrude from the centre of the nucleosome core particle. There is a conserved phosphorylation motif in the C-terminal tail of H2A in lower eukaryotes and of the H2A variant H2AX in higher

eukaryotes, but the C-terminal tail of H2AX is longer than that of the H2A protein shown here. Some of the histone residues that are modified in DNA-damage responses are listed. (Panel adapted, with permission, from ref. 97.) **b**, Crystal structure of the nucleosome core particle. The DNA double helix is shown in green and gold. The four histone proteins are shown in blue (H2A), green (H2B), yellow (H3) and pink (H4). Residues in the nucleosome core particle that are modified and have a role in DNA-damage responses are indicated. (Image reproduced, with permission, from ref. 98.)

Table 1 | Covalent histone modifications that influence DNA-damage responses

Histone	<i>S. cerevisiae</i>			<i>S. pombe</i>			Mammals		
	Amino acid	Modification	Enzyme	Amino acid	Modification	Enzyme	Amino acid	Modification	Enzyme
H2A	S129	Phosphorylation	Mec1 and Tel1	S129	Phosphorylation	Rad3	-	-	-
H2AX	-	-	-	-	-	-	S139	Phosphorylation	ATM, ATR and DNA-PK
H2B	S10	Phosphorylation	Ste20	ND	ND	ND	S14	Phosphorylation	ND
	K123	Ubiquitylation	Rad6 and Bre1	ND	ND	ND	K120	Ubiquitylation	RAD6
H3	K56	Acetylation	Gcn5	ND	ND	ND	ND	ND	ND
	K79	Methylation	Dot1	ND	ND	ND	K79	Methylation	DOT1L
H4	S1	Phosphorylation	CK2	ND	ND	ND	ND	ND	ND
	K5	Acetylation	Esa1	ND	ND	ND	K5	Acetylation	TIP60
	K8	Acetylation	Esa1	ND	ND	ND	K8	Acetylation	TIP60
	K12	Acetylation	Esa1 and Hat1	ND	ND	ND	K12	Acetylation	TIP60
	K16	Acetylation	Esa1	ND	ND	ND	K16	Acetylation	TIP60 and MOF
	ND	ND	ND	K20	Methylation	Set9	K20	Di- or trimethylation	PR-SET7

ND, not determined.

that initially activates the PIKK ATM¹². The observed chromatin decondensation that occurs within seconds of DSB induction is confined to the break site^{6,25}. Chromatin alterations have also been suggested to spread globally, demonstrated most recently by the discovery of an ATM-dependent pathway that mediates large-scale chromatin unfolding²⁶. This finding suggests that ATM mediates a large-scale reconfiguration of chromatin that facilitates the repair of DSBs.

Energy-dependent chromatin remodelling is carried out by large multisubunit protein complexes. The catalytic subunits of these complexes are ATPases of the superfamily 2 group of helicases and are classified into 24 subfamilies²⁷. In this section, we discuss several of these complexes.

RSC

The catalytic subunit of the ATP-dependent chromatin-remodelling complex RSC (remodels the structure of chromatin) is encoded by the essential gene *STH1* in *S. cerevisiae*. RSC can mediate nucleosome sliding, alter histone–DNA contacts and remove histones from DNA²⁷. *In vivo*, the chromatin-remodelling activity of RSC is important for transcriptional regulation of genes that are involved in stress responses and cell-cycle progression²⁸. It seems that RSC can be present in two distinct subcomplexes, containing either Rsc1 or Rsc2, two related non-catalytic proteins with multiple bromodomains²⁹.

Several reports have implicated RSC in DNA-damage responses. Strains lacking *RSC1* were isolated in a screen for sensitivity to ionizing radiation³⁰, and *rsc8* and *rsc30* mutant strains were isolated in a screen for genes required for NHEJ³¹. Interestingly, another report found that *rsc1*, *rsc2* and *sth1* mutant strains were defective in homologous recombination but not NHEJ, apparently as a result of a defect in one of the final stages of homologous recombination (postsynaptic ligation)³². The Sth1 subunit of RSC associates with DSBs *in vivo*, suggesting that it has a direct role in DSB repair^{31,32}. Notably, recent work demonstrated that nucleosomes near a DSB are rapidly mobilized to new positions and that this activity depends on RSC⁵. In addition, in the absence of RSC in *S. cerevisiae*, phosphorylation of H2A (another early response) is delayed⁵. Together, these studies suggest that RSC has a key role at multiple steps during the detection and repair of DSBs.

SWI/SNF

The multisubunit complex SWI/SNF is a founding member of the ATP-dependent chromatin-remodelling activities and has been implicated in DNA-damage responses³². Some *S. cerevisiae* strains with mutations in non-essential subunits of SWI/SNF are sensitive to DNA-damaging agents. RSC and SWI/SNF have significant sequence similarity in their catalytic subunits²⁷. In addition, similarly to RSC, SWI/SNF is implicated in activation of transcription, sliding of nucleosomes, alteration of histone–DNA contacts and removal of histones from DNA²⁷.

SWI/SNF localizes to DSBs, although at a much later time than RSC³² (Fig. 3). In *S. cerevisiae* with mutations in SWI/SNF, synapsis between the broken DNA and homologous donor template is defective, and accumulation of the repair proteins Rad51 and Rad52 on the donor sequence is lost, although their accumulation at the DSB is normal³². These data suggest that SWI/SNF might facilitate remodelling of the donor template to achieve synapsis. Notably, however, *S. cerevisiae* strains with mutations in SWI/SNF have a relatively weak defect in a plasmid-based homologous-recombination assay³² and do not show marked hypersensitivity to DNA-damaging agents in all strain backgrounds. One explanation for this discrepancy is that SWI/SNF is required only for synapsis with donors assembled into heterochromatic (transcriptionally silent) structures. Interestingly, the presence in chromatin of linker histones, which impede homologous-recombination activity *in vivo*³³, inhibits the remodelling activity of SWI/SNF *in vitro*³⁴, suggesting that linker-histone association at donor sequences might reduce the ability of SWI/SNF to facilitate synapsis.

SWI/SNF facilitates DSB repair in mammals also³⁵. Whether SWI/SNF mediates synapsis by remodelling donor sequences has not been investigated, but γ-H2AX formation is defective in cell lines lacking wild-type SWI/SNF activity, as has been observed in *S. cerevisiae* lacking wild-type RSC⁵. Given the sequence similarity between catalytic subunits of RSC and SWI/SNF, it is possible that the function of mediating remodelling at a DSB has shifted between these related complexes during evolution. Alternatively, both complexes might remodel nucleosomes at a DSB in certain conditions.

INO80

The INO80 complex is a large multisubunit complex that has been shown to reposition nucleosomes *in vitro*²⁷ and is important for the transcriptional regulation of certain genes, including those involved in inositol metabolism³⁶. The catalytic subunit of the INO80 complex, Ino80, belongs to the INO80 subfamily of ATP-dependent chromatin-remodelling enzymes²⁷. Unlike other ATP-dependent chromatin-remodelling complexes studied so far, the INO80 complex has helicase activity *in vitro*³⁷.

Similarly to the case for RSC and SWI/SNF, *S. cerevisiae* strains with mutations in non-essential subunits of the INO80 complex are sensitive to DNA-damaging agents³⁷, and the complex also associates with DSBs^{36,38,39}, depending on the phosphorylation status of histone H2A (Fig. 3) (discussed later). However, because the INO80 complex appears at a DSB much later than the very rapid phosphorylation of H2A after DNA damage, and because both factors show distinct localization patterns at the DSB, there must be additional mechanisms that dictate the recruitment of the INO80 complex.

INO80 complexes are enriched at sites of DSBs at late time points after DNA damage^{36,38,39}, suggesting that the INO80 complex is not required

for the initial detection or signalling of the DSB but for the subsequent repair. There is some evidence to support this idea: a lack of ssDNA formation in the absence of the INO80 complex in one study³⁶, and a lack of Rad51 accumulation despite normal RPA binding in another⁷. Interestingly, the INO80 complex has recently been found to contribute little to recombinational repair. Instead, it seems to function in checkpoint adaptation, which is a mechanism that confers on cells the ability to re-enter the cell cycle and divide despite persistent DNA damage. The INO80 complex seems to impinge on this process by affecting the extent of H2A phosphorylation⁴⁰.

SWR

SWR is a complex with the unusual property of removing H2A–H2B dimers from the nucleosome and replacing them with dimers containing H2B and the H2AZ-type variant Htz1 in *S. cerevisiae*⁴¹. The catalytic subunit of SWR, Swr1, is the founding member of its own subfamily of chromatin-remodelling enzymes²⁷.

S. cerevisiae strains lacking HTZ1 or genes encoding subunits of SWR are sensitive to DNA-damaging agents, and SWR has been shown to interact with phosphorylated H2A *in vitro*³⁸. These data raise the intriguing possibility that SWR might remove the phosphorylated H2A

at sites of DSBs and replace it with Htz1. In higher eukaryotes, Swr1 homologues are found in a complex known as TIP60. This complex also contains homologues of the *S. cerevisiae* NuA4 complex, which has HAT activity and a key role in DNA-damage responses (discussed later).

In *D. melanogaster*, the histone variant H2AV is a fusion of an H2AZ-type histone and an H2AX-type histone, and it contains the C-terminal H2AX-type motif that is phosphorylated in response to DNA damage. Interestingly, it was recently found that *D. melanogaster* TIP60 preferentially acetylates nucleosomal phosphorylated H2AV and exchanges it for unmodified H2AV⁴². This study supports the hypothesis that the activities of SWR and NuA4 are coupled in DNA-damage responses and that the early acetylation of histones at a DSB might facilitate the downstream exchange of phosphorylated histone H2A for H2AZ.

The histone code for DSB repair

So how do covalent modifications of histones influence DSB repair?

H2A

The C-terminal tail of H2A and H2AX (discussed later) is phosphorylated in the region of the nucleosome where the DNA enters and exits (Fig. 1a). In condensed structures, such as the 30-nm fibre (also known as the solenoid),

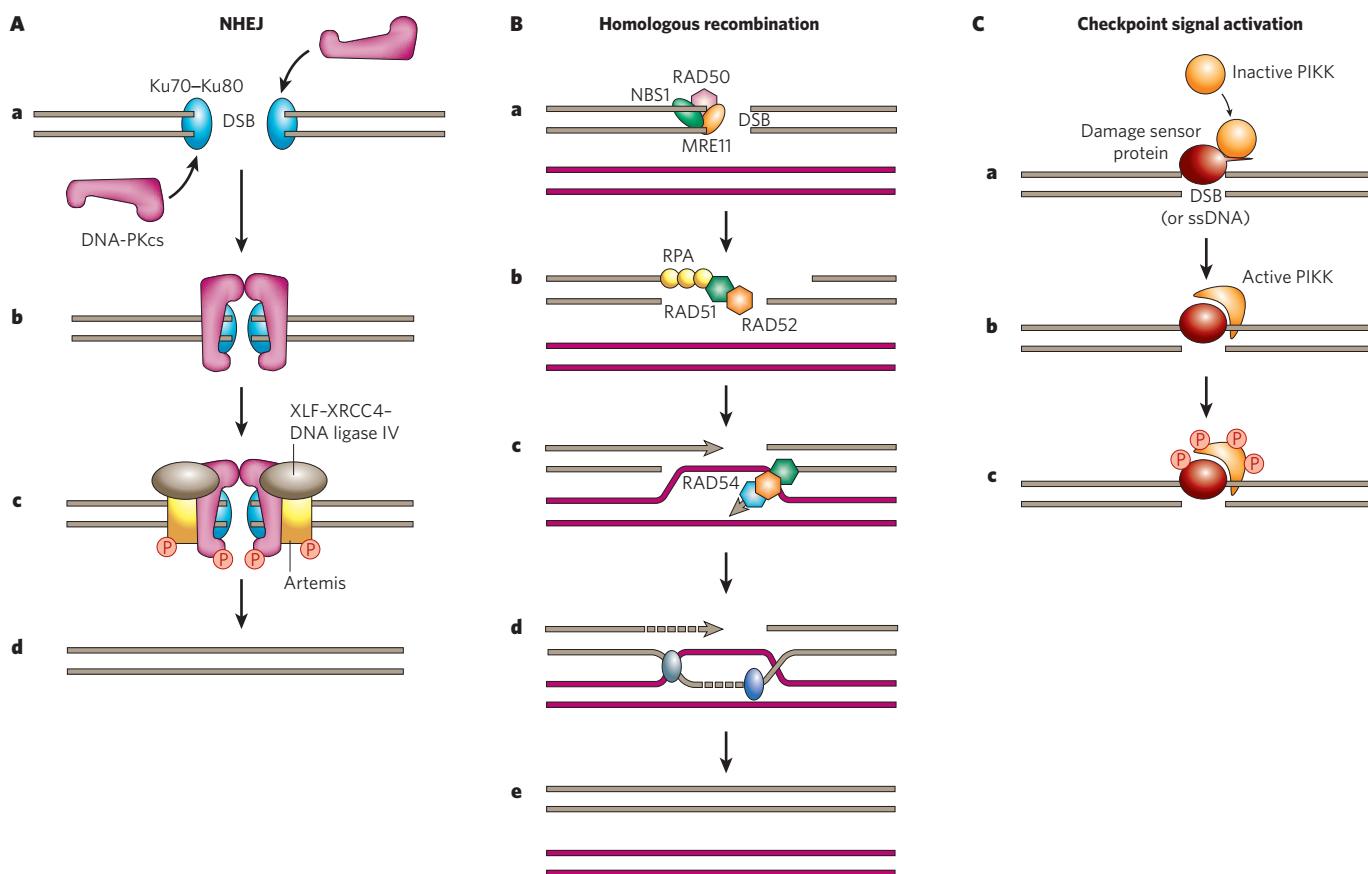


Figure 2 | Main DSB-repair and DSB-induced signalling pathways. There are two main pathways of DNA repair: NHEJ and homologous recombination. **A, NHEJ.** **a**, A DSB is recognized by the Ku dimer (Ku70–Ku80) and DNA-PKcs. **b**, The two DNA ends are synapsed. **c**, DNA-PKcs and Artemis are phosphorylated, and the DNA ends are processed by a complex consisting of XLF (XRCC4-like complex; also known as NHEJ1), XRCC4 (X-ray-repair cross-complementing protein 4) and DNA ligase IV, and by Artemis. **d**, The DNA ends are ligated by DNA ligase IV, and the DNA-repair factors dissociate. **B, Homologous recombination.** **a**, A DSB is recognized by MRN, a complex consisting of MRE11, RAD50 and NBS1. **b**, The DNA ends are resected, allowing binding of RPA, RAD51 and RAD52 to the ssDNA tails. **c**, The ssDNA–protein filament searches for a homologous region in an intact DNA duplex (pink) and initiates pairing, which is facilitated by RAD54. **d**, DNA synthesis occurs from the invading end of the damaged DNA,

extending the repair region and forming a Holliday junction (a cross-stranded structure that occurs between four strands of DNA during recombination). This junction translocates along the DNA in a process mediated by a branch migration complex (purple) and is cleaved by a resolvase (grey). **e**, The DNA ends are rejoined to yield intact duplex products. **C, Checkpoint signal activation.** **a**, DNA lesions (both DSBs and ssDNA) are recognized by a DNA-damage sensor protein (for example, Ku70–Ku80, MRN, or RPA and TOPBP1; brown). An inactive checkpoint kinase (for example, DNA-PKcs, ATM or ATR, respectively; orange) then associates with the sensor. **b**, The protein kinase activity of the checkpoint kinase is stimulated. **c**, Consequently, the checkpoint kinase phosphorylates itself and downstream targets. This signalling leads to triggering of DNA-damage checkpoints. There is an alternative model of checkpoint signal activation, which posits that ATM is activated before recruitment to sites of DNA damage¹².

this region of the nucleosome is at the centre of the fibre, where the linker histone is bound. For this motif to be efficiently phosphorylated, decondensation of the chromatin might be required^{5,6}. Notably, although the phosphorylation of H2A or H2AX is not lost in the absence of RSC- or SWI/SNF-mediated remodelling, respectively^{5,35}, it is delayed or reduced relative to that in wild-type cells. This suggests that chromatin remodelling might either facilitate the exposure of this motif to the PIKKs or allow the efficient translocation of the PIKKs along the flanking chromatin to phosphorylate large numbers of H2A or H2AX molecules.

The phosphorylation of H2A in *S. cerevisiae* seems to be important for several subsequent events. These events include interactions with the INO80, SWR and NuA4 complexes^{36,38,39}, and retention of each of these complexes at DSBs^{36,39}. Furthermore, the phosphorylation of H2A cooperates with the methylation of histone H3 in *S. cerevisiae* or histone H4 in *S. pombe* to mediate the recruitment of Rad9 or its homologue Crb2, respectively^{43–45} (discussed later).

In the absence of phosphorylatable H2A, *S. cerevisiae* cannot achieve wild-type levels of NHEJ activity⁴⁶. In addition, mutation of the serine residue at position 129 (S129) of H2A (H2AS129), when combined with a mutation that affects either NHEJ (*yku70* and/or *yku80*) or homologous recombination (*rad52*) results in higher sensitivity to DNA damage than in *S. cerevisiae* with only one of these mutations, suggesting that phosphorylation of H2A contributes to both DSB-repair pathways⁴⁶. This idea is consistent with studies showing that γ-H2AX is important for the assembly of cohesin (the protein complex responsible for tethering sister chromatids together) at DSBs⁴⁷ and that sister-chromatid recombination defects are similar in cohesin and H2AX mutant strains^{47–49}.

H2A is also modified by acetylation of its N-terminal tail, and this is mediated by NuA4. Acetylation of H2A seems to be important for the ability of cells to survive after DNA damage⁵⁰. In addition, two other phosphorylatable residues in the C-terminal tail of H2A — S122 and T126 — have been implicated in DNA-damage responses^{51,52}, but the precise roles of each residue are not yet known.

H2B

The N-terminal tail of H2B is phosphorylated by the protein kinase Ste20 in response to treatment with hydrogen peroxide, and this is important for the ability of *S. cerevisiae* to undergo apoptosis-like cell death⁵³ (Table 1). H2A is phosphorylated in the same conditions, and this precedes phosphorylation of H2BS10, raising the possibility that this is analogous to the phosphorylation of H2BS14 at DSBs in higher eukaryotes⁵⁴ (Table 1). More recently, phosphorylation of this residue was detected during meiosis in *S. cerevisiae*⁵⁵ and at the immunoglobulin heavy-chain locus in B cells undergoing class-switch recombination and somatic hypermutation⁵⁶.

H3 and H4

The acetylation of conserved lysine residues in the N-terminal tails of H3 and H4 is important for DNA-damage responses^{38,50,57,58} (Table 1). The HATs NuA4, Hat1 and Gcn5 have all been found to contribute to these DNA-damage-dependent modifications. Defects in both NHEJ and homologous recombination have been detected in strains with mutations in either the HATs or the target sites of acetylation^{50,57,58}. Although it is difficult to integrate kinetic data from independent studies, all three HATs seem to appear at sites of DSBs after the appearance of RSC and the phosphorylation of H2A, and before the appearance of the INO80 complex.

In mammalian cells, the TIP60 HAT complex and the HAT cofactor TRRAP (transformation/transcription-domain-associated protein) are recruited to sites of DSBs, where they induce the acetylation of H4 and facilitate homologous recombination⁵⁹ (Table 1). Similarly, the HAT MOF (also known as MYST1) contributes to irradiation-induced acetylation of H4K16 (ref. 60; Table 1), and defects in the acetylation of H3 and H4 have been linked to sensitivity to ionizing radiation and to defective cell-cycle checkpoints^{60,61}.

Several reports have implicated histone deacetylases (HDACs) in DNA-damage responses^{58,62,63}, with HDACs seeming to function late in

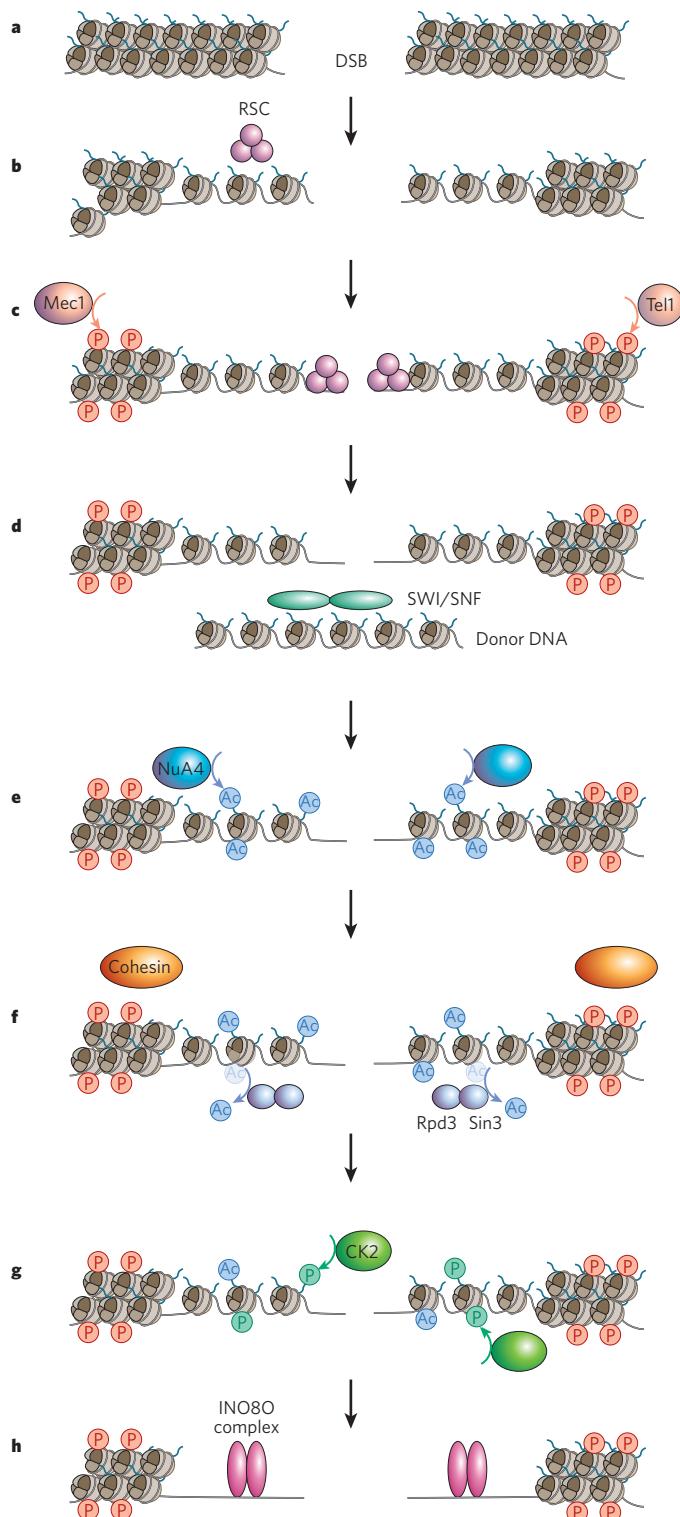


Figure 3 | Early chromatin remodelling and modification events at *S. cerevisiae* DSBs. **a**, A DSB occurs. **b**, RSC then remodels the adjacent chromatin. **c**, The PIKKs Mec1 and Tel1 phosphorylate H2A, and RSC accumulates in the regions flanking the DSB. **d**, SWI/SNF is recruited, and it remodels the donor template chromatin. **e**, The NuA4 complex, which has HAT activity, is then recruited and acetylates the tails of H2A and H4. **f**, Phosphorylated H2A recruits cohesin, which helps to bridge interactions between sister chromatids. **g**, The HDAC complex Sin3–Rpd3 is recruited and removes acetylation from H4. The protein kinase CK2 is also recruited, and this phosphorylates H4. **h**, Finally, the INO80 complex enters the region of the DSB and removes some nucleosomes. This complex is an antagonist of SWR, which replaces core H2A proteins with H2A variants, thereby promoting checkpoint adaptation.

the DNA-repair process. If the acetylation of histones in the vicinity of DNA damage facilitates the access of repair factors to the DNA lesions, then the role of HDACs might be to restore the chromatin to its original state and to terminate both the repair process and any associated checkpoint activity.

H4S1 is phosphorylated by the complex casein kinase II (CK2) in response to DNA damage^{63,64} (Table 1). CK2 associates with the HDAC complex Sin3–Rpd3 (ref. 63), which facilitates late events in the repair process (Fig. 3). Interestingly, phosphorylation of H4S1 inhibits the ability of NuA4 to acetylate the N-terminal tail of H4, suggesting a mechanism by which the chromatin modifications are re-established. In other words, after the NuA4-dependent acetylation is removed by the HDAC Sin3–Rpd3, the associated CK2 phosphorylates H4S1, thus preventing reacetylation by NuA4 and turning off DNA-damage signals.

The histone code for DSB-induced checkpoints

So how do covalent modifications of histones facilitate signalling of DNA-damage checkpoints?

H2AX

H2ax-knockout mouse cells are defective in the triggering of the G2–M-boundary checkpoint in response to low doses of ionizing radiation⁶⁵. We have suggested that H2AX-mediated concentration of proteins into DSB-flanking chromatin is crucial for amplifying the DSB signal when the damage load is below a certain threshold⁶⁵. In addition to its role in the induction of the DNA-damage checkpoint, phosphorylation of H2AX seems to function in checkpoint maintenance in *S. cerevisiae*. This process involves a complex containing the phosphatase Pph3, which regulates the dephosphorylation of H2AX⁶⁶. Pph3-mediated reversion of γ-H2AX to its unphosphorylated state triggers release from the damage checkpoint, possibly because it disrupts the association between checkpoint proteins and the chromatin surrounding the damage site^{66,67}. Consistent with this finding, H2AX deficiency in mice leads to the premature dissociation of factors from DSBs²².

H2AX is also required for efficient resolution of DSBs during class-switch recombination in lymphocytes^{23,24}. This is possibly because the spreading of the γ-H2AX chromatin modification across the domain involved in class switching facilitates the synapsis between DSBs that are separated by up to 200 kilobases^{24,68–71}. Class switching is initiated by AID⁷², an enzyme that produces programmed DSBs at the immunoglobulin locus switch regions²³. In the absence of H2AX, these programmed DSBs fail to be joined, resulting in frequent chromosomal abnormalities involving the immunoglobulin locus^{73,74}.

H3

Acetylation of H3K56, which is just upstream of the histone fold domain of H3 (Fig. 1b), is important for mediating DNA-damage responses^{75,76}. In *S. cerevisiae*, this modification is present on newly synthesized histones^{76,77} and is maintained in a Rad9- and Mec1-dependent manner if DSBs are encountered⁷⁶. Acetylation of H3K56 seems to weaken interactions between H3 and DNA, thereby promoting chromatin unfolding and increasing the efficiency of damage-induced signalling during DNA replication. Acetylation is mediated by the HAT Rtt109, together with the histone chaperone Asf1 (refs 78, 79). This DNA-damage-induced checkpoint regulation of H3K56 is mediated by the HDACs Hst3 and Hst4 (refs 80, 81).

Although H3 is methylated at several sites, only methylation of K79, which is in the histone fold domain (Fig. 1b), by the *S. cerevisiae* methyltransferase Dot1 (Table 1) seems to have a role in the DNA-damage checkpoint. This modification depends on prior ubiquitylation of histone H2BK123 by the ubiquitin–protein ligase complex Rad6–Brel–Lge1. Mutation of genes involved in this pathway (such as *DOT1* or *BRE1*) or mutations in the modified histone residues (H2BK123 or H3K79) cause defects in the G1- and intra-S-phase checkpoints but not in the G2–M-boundary checkpoint^{82,83}. Interestingly, these mutant strains are not particularly sensitive to DNA-damaging agents. Methylation of H3K79 affects DNA-damage control in mammals because this modification mediates

the rapid initial recruitment of 53BP1 (p53-binding protein 1; also known as TP53BP1)^{65,84,85} — the mammalian homologue of *S. cerevisiae* Rad9 and *S. pombe* Crb2 — through interactions with the tudor domain of 53BP1 (ref. 8; Table 1). 53BP1 localizes to sites of DSBs and interacts with γ-H2AX. Similarly to H2AX, 53BP1 is required for the repair of DSBs produced during class-switch recombination in lymphocytes; indeed, deficiency in 53BP1 has a more severe phenotype, resulting in almost complete abrogation of the reaction^{68–70}. Unlike phosphorylation of the H2AX C-terminal tail, methylation of H3K79 is not induced in response to DNA damage, leading to the idea that the accessibility of constitutively methylated H3K79 increases at sites of DSBs⁸. One possibility for how this might occur is that chromatin relaxation induced by either ATP-dependent remodelling enzymes or the DSB itself exposes methylated H3K79, facilitating the recruitment of 53BP1 to DSBs^{68,86}. Phosphorylated H2AX subsequently stabilizes this interaction on flanking chromatin. A similar mechanism that combines ATP-dependent chromatin-remodelling activities, histone tail phosphorylation and static histone lysine methylation might regulate the binding and retention of other proteins involved in the DNA-damage response.

H4

In *S. pombe*, the accumulation of Crb2 into foci is independently regulated by the methylation of H4K20 (which is carried out by the histone methyltransferase Set9) and by the phosphorylation of H2A^{87,88} (Table 1). Whereas *crb2*-knockout strains have a severely defective DNA-damage response, strains with mutations in Set9 or with amino-acid substitutions at specific positions in H2A or H4 (H2AS129A and H4K20R) are only moderately sensitive to DNA damage and have mild impairment in DNA-damage checkpoints. More specifically, although the mutant cells seem to be fully capable of initially triggering the checkpoint arrest, they are less able to maintain the DNA-damage checkpoint^{87,88}. Thus, although the formation of Crb2 foci and the maintenance of checkpoints depends on both H4K20 methylation and H2AS129 phosphorylation, these modifications contribute only partly to the Crb2-dependent checkpoint.

In mammals, methylation of H3K79 is restricted to euchromatic regions (which are relatively decondensed and gene-rich), whereas trimethylation of H4K20 is enriched in pericentromeric heterochromatin (which is highly condensed and gene-poor)⁸⁹. Although methylated H3K79 might be the main target of 53BP1, H4K20 trimethylation might also contribute to 53BP1 recruitment, depending on whether the DSB occurs in euchromatin or heterochromatin (Table 1).

Distinct histone methyltransferases methylate H4K20; the PR-SET7 (also known as SET8) enzyme shows a preference for carrying out monomethylation, whereas SUV4-20H1 and SUV4-20H2 mediate dimethylation and trimethylation, respectively. So far, SUV4-20H1 and SUV4-20H2 have not been implicated in the mammalian DNA-damage response. Recent experiments suggest that 53BP1 binds dimethylated H4K20 and that PR-SET7 is required for the irradiation-induced formation of 53BP1 foci⁹⁰. However, further genetic studies are needed to determine the role of dimethylated H4K20 in targeting 53BP1 to DSBs.

Reading the DNA-damage histone code

How are histone-modifying enzymes and ATP-dependent chromatin remodelers targeted to specific loci? During transcription, the targeting does not seem to occur through direct binding of the underlying DNA sequence. Instead, interactions between histone-modifying enzymes and sequence-specific transcription factors seem to be necessary. For example, SWI/SNF, which remodels chromatin, and HAT complexes are recruited to promoters by direct interaction with transcriptional activators^{91,92}. After a modification pattern has been established, it is ‘read’ by chromatin-remodelling factors (effectors) that increase the efficiency of transcription either by altering the position or mobility of nucleosomes or by promoting the eviction of histones.

Just as histone marks are ‘delivered’ to specific promoters by transcription factors, histone-modifying enzymes and chromatin-remodelling enzymes might be targeted to DSBs by associating with DSB sensor proteins. One clear example that supports this mechanism is phosphorylation

of H2AX by the PIKKs ATM and DNA-PKcs. The prevailing view is that the DSB is initially detected by MRN or Ku70–Ku80, which, similarly to sequence-specific transcription factors, have high affinities for DSBs (K_d of $(1.5\text{--}2) \times 10^{-11}$ M for the Ku complex at physiological salt concentrations)⁹³. After the PIKKs have been guided to the DSBs by interactions with the DSB sensor proteins, the PIKKs phosphorylate H2AX. The DSB effector protein MDC1, which is recruited to the break site independently of H2AX, recognizes the phosphorylation of H2AX through a low-affinity interaction (K_d of 2.2×10^{-6} M)⁹⁴. This promotes the further accumulation and stabilization of both proteins involved in the DNA-damage response and chromatin-remodelling factors in the chromatin environment surrounding the break^{94,95}. Generally, chromatin-remodelling factors contain domains (for example, bromodomains and chromodomains) that bind to modified histones with low affinity (K_d of 10^{-6} to 10^{-4} M). Therefore, histone marks might not be the main gene-targeting mechanism used by chromatin-remodelling enzymes involved in transcription or DSB repair, but these marks might be required for stabilizing the interaction of these enzymes with flanking chromatin.

Whereas active promoters (which are generally present in euchromatic regions) are accessible to sequence-specific transcription factors, heterochromatic DNA (which is highly compact) is inaccessible to the transcription machinery. Similarly, a DSB in euchromatin might be directly detected by a DSB sensor protein, whereas the local environment surrounding damage in heterochromatin might need to be reconfigured. The findings that a DSB induces a rapid local decrease in the density of the chromatin fibre⁶ and that nucleosomes in the vicinity of the DSB are repositioned⁵ support the idea that ATP-dependent chromatin-remodelling factors have an early role in the DNA-damage response. Local relaxation of the chromatin structure as a result of DSBs might also be an important mechanism that drives transcription, because various promoters require topoisomerase-IIβ-mediated DSB formation for gene activation⁹⁶.

Complexity of the DNA-damage histone code

Numerous histone modifications influence gene expression, and the combination, the timing and the order of appearance of these modifications vary at individual promoters. By analogy, it is unlikely that all DSBs follow the same pattern of chromatin modulation during repair. For example, it is possible that distinct chromatin intermediates are generated if a DSB is encountered in G1 phase as opposed to at the S–G2 boundary. The nature of the events that occur during homologous recombination might dictate a greater need for chromatin structure to be opened up, because large stretches of DNA are required for steps such as strand invasion. Indeed, this is highlighted by the finding that SWI/SNF seems to be important for the accumulation of Rad51 and Rad52 on the donor sequence in *S. cerevisiae*, whereas the accumulation of these proteins at the DSB is normal in *S. cerevisiae* with mutations in SWI/SNF³². Similarly, drastic decondensation events such as histone eviction near the DNA ends might occur in G2 phase when homologous recombination is preferentially used. By contrast, successful NHEJ does not require homologous donor sequences and requires far less exposure of DNA surrounding the DSB. Consequently, perhaps less decondensation of the chromatin that flanks the lesion is needed.

The induction and maintenance of the damage checkpoint signal might also require distinct patterns of chromatin alterations. To trigger a checkpoint, the lesion must be recognized by PIKKs and associated effectors, and the unwrapping or sliding of nucleosomes might make the lesion more accessible. The long-term maintenance of the relaxed chromatin configuration might involve modifications, such as phosphorylation of H2AX, that promote the sustained interactions between DNA-damage-response proteins and a persistent lesion²⁵.

It is plausible that the recruitment and activation of chromatin-modifying machinery is also strictly regulated, perhaps by cyclin-dependent kinase activity and/or by PIKK-mediated phosphorylation events. In addition to the chromatin-modifying machinery being regulated in a cell-cycle-specific manner, the ‘marks’ that remain at the site of a DSB could be read differently depending on the stage of the cell cycle. In

this regard, it is interesting that H2A phosphorylation (which occurs in response to DSBs regardless of the phase of the cell cycle) works together with H3 methylation (which is constitutively present) specifically in G1 phase to mediate checkpoint responses^{43,45}.

Concluding remarks

The phosphorylation of H2AX was the first chromatin-associated event discovered to occur at DSBs and was the only known event for several years. Today, most of the chromatin machinery that was initially identified in transcriptional studies has been found to have a role in the DNA-damage response. A major challenge will be to determine more precisely how the activities that remodel chromatin structure are integrated with the events associated with DNA-damage repair. To accomplish this, we need to uncover the molecular determinants of the recruitment of chromatin remodellers to DNA lesions, to define how histone-modifying enzymes and chromatin-remodelling enzymes interact with core components of the DNA-damage surveillance and repair machinery, and to characterize mechanistically how chromatin modifications influence specific biochemical steps in the DNA-damage response. This basic research quest ultimately holds promise for improved diagnostic, predictive and therapeutic measures for treating disorders associated with genomic instability. ■

1. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
2. Turner, B. M. Histone acetylation and an epigenetic code. *Bioessays* **22**, 836–845 (2000).
3. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844–847 (2006).
4. Takahashi, K. & Kaneko, I. Changes in nuclease sensitivity of mammalian cells after irradiation with ^{60}Co γ -rays. *Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med.* **48**, 389–395 (1985).
5. Shim, E. Y. *et al.* RSC mobilizes nucleosomes to improve accessibility of repair machinery to the damaged chromatin. *Mol. Cell. Biol.* **27**, 1602–1613 (2007).
6. Kruhlak, M. J. *et al.* Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. *J. Cell Biol.* **172**, 823–834 (2006).
7. Tsukuda, T., Fleming, A. B., Nickoloff, J. A. & Osley, M. A. Chromatin remodelling at a DNA double-strand break site in *Saccharomyces cerevisiae*. *Nature* **438**, 379–383 (2005).
8. Huyen, Y. *et al.* Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406–411 (2004).
9. Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Cell Biol.* **146**, 905–916 (1999).
10. Rouse, J. & Jackson, S. P. Interfaces between the detection, signalling, and repair of DNA damage. *Science* **297**, 547–551 (2002).
11. Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).
12. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
13. Lee, J. H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11–Rad50–Nbs1 complex. *Science* **308**, 551–554 (2005).
14. Dupre, A., Boyer-Chatenet, L. & Gautier, J. Two-step activation of ATM by DNA and the Mre11–Rad50–Nbs1 complex. *Nature Struct. Mol. Biol.* **13**, 451–457 (2006).
15. You, Z., Chahwan, C., Bailis, J., Hunter, T. & Russell, P. ATM activation and its recruitment to damaged DNA require binding to the C terminus of Nbs1. *Mol. Cell. Biol.* **25**, 5363–5379 (2005).
16. Pellegrini, M. *et al.* Autophosphorylation at serine 1987 is dispensable for murine ATM activation *in vivo*. *Nature* **443**, 222–225 (2006).
17. Meek, C. K., Gupta, S., Ramsden D. A. & Lees-Miller, S. P. The DNA-dependent protein kinase: the director at the end. *Immunol. Rev.* **200**, 132–141 (2004).
18. Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).
19. Kumagai, A., Lee, J., Yoo H. Y. & Dunphy, W. G. TopBP1 activates the ATR-ATRIP complex. *Cell* **124**, 943–955 (2006).
20. Fernandez-Capetillo, O., Celeste, A. & Nussenzweig, A. Focusing on foci: H2AX and the recruitment of DNA-damage response factors. *Cell Cycle* **2**, 426–427 (2003).
21. Bekker-Jensen S. *et al.* Spatial organization of the mammalian genome surveillance machinery in response to DNA strand breaks. *J. Cell Biol.* **173**, 195–206 (2006).
22. Celeste, A. *et al.* Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nature Cell Biol.* **5**, 675–679 (2003).
23. Petersen, S. *et al.* AID is required to initiate Nbs1/γ-H2AX focus formation and mutations at sites of class switching. *Nature* **414**, 660–665 (2001).
24. Reina-San-Martin, B. *et al.* H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *J. Exp. Med.* **197**, 1767–1778 (2003).
25. Kruhlak, M. J., Celeste, A. & Nussenzweig, A. Spatio-temporal dynamics of chromatin containing DNA breaks. *Cell Cycle* **5**, 1910–1912 (2006).
26. Ziv, Y. *et al.* Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1-dependent pathway. *Nature Cell Biol.* **8**, 870–876 (2006).
27. Flaus, A., Martin, D. M., Barton G. J. & Owen-Hughes, T. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res.* **34**, 2887–2905 (2006).

28. Angus-Hill, M. L. et al. A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control. *Mol. Cell* **7**, 741–751 (2001).

29. Cairns, B. R. et al. Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and bromodomains. *Mol. Cell* **4**, 715–723 (1999).

30. Bennett, C. B. et al. Genes required for ionizing radiation resistance in yeast. *Nature Genet.* **29**, 426–434 (2001).

31. Shim, E. Y., Ma, J. L., Oum, J. H., Yanez, Y. & Lee, S. E. The yeast chromatin remodeler RSC complex facilitates end joining repair of DNA double-strand breaks. *Mol. Cell. Biol.* **25**, 3934–3944 (2005).

32. Chai, B., Huang, J., Cairns, B. R. & Laurent, B. C. Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair. *Genes Dev.* **19**, 1656–1661 (2005).

33. Downs, J. A., Kosmidou, E., Morgan, A. & Jackson, S. P. Suppression of homologous recombination by the *Saccharomyces cerevisiae* linker histone. *Mol. Cell* **11**, 1685–1692 (2003).

34. Hill, D. A. & Imbalzano, A. N. Human SWI/SNF nucleosome remodeling activity is partially inhibited by linker histone H1. *Biochemistry* **39**, 11649–11656 (2000).

35. Park, J. H. et al. Mammalian SWI/SNF complexes facilitate DNA double-strand break repair by promoting γ-H2AX induction. *EMBO J.* **25**, 3986–3997 (2006).

36. van Attikum, H., Fritsch, O., Hohn, B. & Gasser, S. M. Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. *Cell* **119**, 777–788 (2004).

37. Shen, X., Mizuguchi, G., Hamich, A. & Wu, C. A chromatin remodelling complex involved in transcription and DNA processing. *Nature* **406**, 541–544 (2000).

38. Downs, J. A. et al. Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. *Mol. Cell* **16**, 979–990 (2004).

39. Morrison, A. J. et al. INO80 and γ-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. *Cell* **119**, 767–775 (2004).

40. Papamichos-Chronakis, M., Krebs, J. E. & Peterson, C. L. Interplay between Ino80 and Swr1 chromatin remodeling enzymes regulates cell cycle checkpoint adaptation in response to DNA damage. *Genes Dev.* **20**, 2437–2449 (2006).

41. Mizuguchi, G. et al. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343–348 (2004).

42. Kusch, T. et al. Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. *Science* **306**, 2084–2087 (2004).

43. Toh, G. W. et al. Histone H2A phosphorylation and H3 methylation are required for a novel Rad9 DSB repair function following checkpoint activation. *DNA Repair (Amst.)* **5**, 693–703 (2006).

44. Nakamura, T. M., Du, L.-L., Redon, C. & Russell, P. Histone H2A phosphorylation controls Crb2 recruitment at DNA breaks, maintains checkpoint arrest, and influences DNA repair in fission yeast. *Mol. Cell. Biol.* **24**, 6215–6230 (2004).

45. Javaheri, A. et al. Yeast G1 DNA damage checkpoint regulation by H2A phosphorylation is independent of chromatin remodeling. *Proc. Natl Acad. Sci. USA* **103**, 13771–13776 (2006).

46. Downs, J. A., Lowndes, N. F. & Jackson, S. P. A role for *Saccharomyces cerevisiae* histone H2A in DNA repair. *Nature* **408**, 1001–1004 (2000).

47. Unal, E. et al. DNA damage response pathway uses histone modification to assemble a double-strand-break-specific cohesin domain. *Mol. Cell* **16**, 991–1002 (2004).

48. Strom, L., Lindroos, H. B., Shirahige, K. & Sjogren, C. Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Mol. Cell* **16**, 1003–1015 (2004).

49. Xie, A. et al. Control of sister chromatid recombination by histone H2AX. *Mol. Cell* **16**, 1017–1025 (2004).

50. Bird, A. W. et al. Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. *Nature* **419**, 411–415 (2002).

51. Harvey, A. C., Jackson, S. P. & Downs, J. A. *Saccharomyces cerevisiae* histone H2A Ser122 facilitates DNA repair. *Genetics* **170**, 543–553 (2005).

52. Wyatt, H. R., Liaw, H., Green, G. R. & Lustig, A. J. Multiple roles for *Saccharomyces cerevisiae* histone H2A in telomeric position effect, Spt phenotypes and double-strand-break repair. *Genetics* **164**, 47–64 (2003).

53. Ahn, S. H. et al. Sterile 20 kinase phosphorylates histone H2B at serine 10 during hydrogen peroxide-induced apoptosis in *S. cerevisiae*. *Cell* **120**, 25–36 (2005).

54. Fernandez-Capetillo, O., Allis, C. D. & Nussenzweig, A. Phosphorylation of histone H2B at DNA double-strand breaks. *J. Exp. Med.* **199**, 1671–1677 (2004).

55. Ahn, S. H., Henderson, K. A., Keeney, S. & Allis, C. D. H2B (Ser10) phosphorylation is induced during apoptosis and meiosis in *S. cerevisiae*. *Cell Cycle* **4**, 780–783 (2005).

56. Odegaard, V. H., Kim, S. T., Anderson, S. M., Shlomchik, M. J. & Schatz, D. J. Histone modifications associated with somatic hypermutation. *Immunity* **23**, 101–110 (2005).

57. Qin, S. & Parthun, M. R. Recruitment of the type B histone acetyltransferase Hat1p to chromatin is linked to DNA double-strand breaks. *Mol. Cell. Biol.* **26**, 3649–3658 (2006).

58. Tamburini, B. A. & Tyler, J. K. Localized histone acetylation and deacetylation triggered by the homologous recombination pathway of double-strand DNA repair. *Mol. Cell. Biol.* **25**, 4903–4913 (2005).

59. Murr, R. et al. Histone acetylation by Trapp-Tip60 modulates loading of repair proteins and repair of DNA double-strand breaks. *Nature Cell Biol.* **8**, 91–99 (2006).

60. Gupta, A. et al. Involvement of human MOF in ATM function. *Mol. Cell. Biol.* **25**, 5292–5305 (2005).

61. Birger, Y. et al. Increased tumorigenicity and sensitivity to ionizing radiation upon loss of chromosomal protein HMGN1. *Cancer Res.* **65**, 6711–6718 (2005).

62. Jazayeri, A., McAinsh, A. D. & Jackson, S. P. *Saccharomyces cerevisiae* Sin3p facilitates DNA double-strand break repair. *Proc. Natl Acad. Sci. USA* **101**, 1644–1649 (2004).

63. Utley, R., Lacoste, N., Jobin-Robitaillé, O., Allard, S. & Côté, J. Regulation of NuA4 histone acetyltransferase activity in transcription and DNA repair by phosphorylation of histone H4. *Mol. Cell. Biol.* **25**, 8179–8190 (2005).

64. Cheung, W. L. et al. Phosphorylation of histone H4 serine 1 during DNA damage requires casein kinase II in *S. cerevisiae*. *Curr. Biol.* **15**, 656–660 (2005).

65. Fernandez-Capetillo, O. et al. DNA damage-induced G2-M checkpoint activation by histone H2AX and 53BP1. *Nature Cell Biol.* **4**, 993–997 (2002).

66. Keogh, M. C. et al. A phosphatase complex that dephosphorylates γH2AX regulates DNA damage checkpoint recovery. *Nature* **439**, 497–501 (2006).

67. Nussenzweig, A. & Paull, T. DNA repair: tails of histones lost. *Nature* **439**, 406–407 (2006).

68. Reina-San-Martin, B., Chen, J., Nussenzweig, A. & Nussenzweig, M. C. Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP1^{-/-} B cells. *Eur. J. Immunol.* **37**, 235–239 (2007).

69. Manis, J. P. et al. 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. *Nature Immunol.* **5**, 481–487 (2004).

70. Ward, I. M. et al. 53BP1 is required for class switch recombination. *J. Cell. Biol.* **165**, 459–464 (2004).

71. Reina-San-Martin, B., Nussenzweig, M. C., Nussenzweig, A. & Difilippantonio, S. Genomic instability, endoreduplication, and diminished Ig class-switch recombination in B cells lacking Nbs1. *Proc. Natl Acad. Sci. USA* **102**, 1590–1595 (2005).

72. Honjo, T., Nagaoka, H., Shinkura, R. & Muramatsu, M. AID to overcome the limitations of genomic information. *Nature Immunol.* **6**, 655–661 (2005).

73. Franco, S. et al. H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol. Cell* **21**, 201–214 (2006).

74. Ramiro, A. R. et al. Role of genomic instability and p53 in AID-induced c-myc-lgh translocations. *Nature* **440**, 105–109 (2006).

75. Ozdemir, A. et al. Characterization of lysine 56 of histone H3 as an acetylation site in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**, 25949–25952 (2005).

76. Masumoto, H., Hawke, D., Kobayashi, R. & Verreault, A. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436**, 294–298 (2005).

77. Dirksen, E. H. et al. Human lymphoblastoid proteome analysis reveals a role for the inhibitor of acetyltransferases complex in DNA double-strand break response. *Cancer Res.* **66**, 1473–1480 (2006).

78. Driscoll, R., Hudson, A. & Jackson, S. P. Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56. *Science* **315**, 649–652 (2007).

79. Collins, S. R. et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).

80. Maas, N. L., Miller, K. M., DeFazio, L. G. & Toczkowski, D. P. Cell cycle and checkpoint regulation of histone H3 K56 acetylation by Hst3 and Hst4. *Mol. Cell* **23**, 109–119 (2006).

81. Celic, I. H. et al. The sirtuins Hst3 and Hst4p preserve genome integrity by controlling histone H3 lysine 56 deacetylation. *Curr. Biol.* **16**, 1280–1289 (2006).

82. Wysocki, R. et al. Role of Dot1-dependent histone H3 methylation in G1 and S phase DNA damage checkpoint functions of Rad9. *Mol. Cell. Biol.* **25**, 8430–8443 (2005).

83. Giannattasio, M., Lazzaro, F., Plevani, P. & Muzi-Falconi, M. The DNA damage checkpoint response requires histone H2B ubiquitination by Rad6-Bre1 and H3 methylation by Dot1. *J. Biol. Chem.* **280**, 9879–9886 (2005).

84. DiTullio, R. A. Jr. et al. 53BP1 functions in an ATM-dependent checkpoint pathway that is constitutively activated in human cancer. *Nature Cell Biol.* **4**, 998–1002 (2002).

85. Cao, R. et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).

86. Bekker-Jensen, S., Lukas, C., Melander, F., Bartek, J. & Lukas, J. Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFBD1. *J. Cell. Biol.* **170**, 201–211 (2005).

87. Sanders, S. L. et al. Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell* **119**, 603–614 (2004).

88. Du, L. L., Nakamura, T. M. & Russell, P. Histone modification-dependent and -independent pathways for recruitment of checkpoint protein Crb2 to double-strand breaks. *Genes Dev.* **20**, 1583–1596 (2006).

89. Schotta, G. et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).

90. Botuyan, M. V. et al. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).

91. Ryan, M. P., Jones, R. & Morse, R. H. SWI-SNF complex participation in transcriptional activation at a step subsequent to activator binding. *Mol. Cell. Biol.* **18**, 1774–1782 (1998).

92. Brown, C. E. et al. Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. *Science* **292**, 2333–2337 (2001).

93. Falzon, M., Fewell, J. W. & Kuff, E. L. EBP-80, a transcription factor closely resembling the human autoantigen Ku, recognizes single- to double-strand transitions in DNA. *J. Biol. Chem.* **268**, 10546–10552 (1993).

94. Stucki, M. et al. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123**, 1213–1226 (2005).

95. Lou, Z. et al. MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Mol. Cell* **21**, 187–200 (2006).

96. Ju, B. G. et al. A topoisomerase IIβ-mediated dsDNA break required for regulated transcription. *Science* **312**, 1798–1802 (2006).

97. Wolffe, A. P. & Hayes, J. J. Chromatin disruption and modification. *Nucleic Acids Res.* **27**, 711–720 (1999).

98. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).

Acknowledgements We apologize to colleagues whose research could not be cited because of space limitations. J.A.D. is a Jenner Fellow of the Lister Institute of Preventive Medicine. M.C.N. is supported by grants from the National Institutes of Health and is a Howard Hughes Medical Institute Investigator. A.N. is supported by the Intramural Research Program of the National Institutes of Health (at the Center for Cancer Research, National Cancer Institute), and the A-T Children's Project.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to J.A.D. (j.a.downs@sussex.ac.uk) or A.N. (andre_nussenzweig@nih.gov).

ARTICLES

Treatment of diabetes and atherosclerosis by inhibiting fatty-acid-binding protein aP2

Masato Furuhashi¹, Gürol Tuncman¹, Cem Z. Görgün¹, Liza Makowski^{1,4}, Genichi Atsumi^{1†}, Eric Vaillancourt¹, Keita Kono¹, Vladimir R. Babaev², Sergio Fazio², MacRae F. Linton², Richard Sulsky³, Jeffrey A. Robl³, Rex A. Parker³ & Gökhan S. Hotamisligil¹

Adipocyte fatty-acid-binding protein, aP2 (FABP4) is expressed in adipocytes and macrophages, and integrates inflammatory and metabolic responses. Studies in aP2-deficient mice have shown that this lipid chaperone has a significant role in several aspects of metabolic syndrome, including type 2 diabetes and atherosclerosis. Here we demonstrate that an orally active small-molecule inhibitor of aP2 is an effective therapeutic agent against severe atherosclerosis and type 2 diabetes in mouse models. In macrophage and adipocyte cell lines with or without aP2, we also show the target specificity of this chemical intervention and its mechanisms of action on metabolic and inflammatory pathways. Our findings demonstrate that targeting aP2 with small-molecule inhibitors is possible and can lead to a new class of powerful therapeutic agents to prevent and treat metabolic diseases such as type 2 diabetes and atherosclerosis.

Lipids and lipid signals are critical in the integration of metabolic and inflammatory response systems and consequently play significant parts in the pathogenesis of a cluster of chronic metabolic diseases, including type 2 diabetes, fatty liver disease and atherosclerosis¹. However, how lipids couple to target signalling pathways or metabolic processes and how their intracellular trafficking is regulated are poorly understood. Cytoplasmic fatty-acid-binding proteins (FABPs) are a family of 14–15-kDa proteins that bind with high affinity to hydrophobic ligands such as saturated and unsaturated long-chain fatty acids and eicosanoids such as hydroxyeicosatetraenoic acid, leukotrienes and prostaglandins². The adipocyte FABP, aP2 (FABP4), is highly expressed in adipocytes and regulated by peroxisome-proliferator-activated receptor- γ (PPAR γ) agonists, insulin and fatty acids^{2–5}.

Studies in aP2-deficient mice have shown that aP2 has a significant role in many aspects of metabolic syndrome. Deficiency of aP2 partially protects mice against the development of insulin resistance associated with genetic or diet-induced obesity^{6,7}. Adipocytes of $aP2^{-/-}$ mice have reduced efficiency of lipid transport *in vitro* and *in vivo*, and yet exhibit only minor changes in serum lipids⁸. Interestingly, recent studies demonstrated that aP2 is also expressed in macrophages and regulated by phorbol 12-myristate 13-acetate, lipopolysaccharide, oxidized low-density lipoproteins and PPAR γ ligands^{9–12}. The macrophage is a critical site of FABP action, and total or macrophage-specific aP2-deficiency leads to a marked protection against early and advanced atherosclerosis in apolipoprotein E-deficient (*Apoe*^{-/-}) mice^{9,13}.

These findings indicate an important role for aP2 in the development of major components of metabolic syndrome through its distinct actions in adipocytes and macrophages of integrating metabolic and inflammatory responses. Hence, pharmacological agents that modify FABP function may offer therapeutic opportunities for many components of metabolic syndrome, such as insulin resistance, type 2 diabetes, and atherosclerosis. Here, we demonstrate the first evidence

of the efficacy of a novel chemical aP2 inhibitor in experimental models.

Inhibition of aP2 in cellular models

BMS309403 (Fig. 1a) is a rationally designed, potent, and selective inhibitor of aP2 that interacts with the fatty-acid-binding pocket within the interior of the protein and competitively inhibits the binding of endogenous fatty acids. In a fluorescent 1,8-anilino-8-naphthalene sulphonate (ANS) binding displacement assay, BMS309403 exhibited K_i values <2 nM for both mouse and human aP2, compared with 250 nM for muscle FABP (FABP3) and 350 nM for mal1 (FABP5)¹⁴. In this assay, the endogenous fatty acids palmitic acid and oleic acid exhibited aP2 K_i values of 336 and 185 nM, respectively. Results of X-ray crystallography studies suggested the specific interactions of BMS309403 with key residues in the fatty-acid-binding pocket are the basis of its high *in vitro* binding affinity and selectivity for aP2 over other FABPs¹⁴.

To test the specificity of aP2 inhibition by BMS309403, we developed and used a cellular system with $aP2^{+/+}$ and $aP2^{-/-}$ macrophage cell lines^{9,15}. In addition, we reconstituted aP2 expression in the $aP2^{-/-}$ cells ($aP2^{-/-}R$). As shown in Fig. 1b, aP2 protein was expressed in the THP-1 (a human monocytic leukaemia cell line), $aP2^{+/+}$, and $aP2^{-/-}R$ macrophages but was not detected in the $aP2^{-/-}$ macrophages. In all of the cell lines, mal1 was present (Fig. 1b, c). Similarly, aP2 messenger RNA was readily detectable in THP-1, $aP2^{+/+}$ and $aP2^{-/-}R$ but not in the $aP2^{-/-}$ macrophages (Fig. 1c). In this system, we examined the impact of aP2 inhibition on production of monocyte chemoattractant protein (MCP)-1 (also known as CCL2), an important aP2-regulated atherogenic product^{9,15}, in macrophages.

Treatment with BMS309403 significantly decreased MCP-1 production from THP-1 macrophages in a dose- and time-dependent manner (Fig. 1d, Supplementary Fig. 1). To address whether this effect is specific, we next investigated MCP-1 production using

¹Department of Genetics and Complex Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ²Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee 37232, USA. ³Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543, USA. ⁴Present address: Department of Medicine, Division of Endocrinology, Metabolism, and Nutrition, Duke University Medical Center, Durham, North Carolina 27704, USA. [†]Present address: Clinical Molecular Biology, Teikyo University, Kanagawa 199-0195, Japan.

aP2^{+/+}, *aP2^{-/-}* or *aP2^{-/-R}* mouse macrophage cell lines. In a similar way to THP-1 cells, production of MCP-1 from macrophages was decreased in *aP2^{+/+}* cells in a dose-dependent fashion. In contrast, BMS309403 had no effect on MCP-1 production in *aP2^{-/-}* cells at any dose tested. However, re-expression of aP2 (*aP2^{-/-R}*) rendered the *aP2^{-/-}* cells responsive to BMS309403 treatment, resulting in a dose-dependent reduction in MCP-1 production and demonstrating the target specificity of this compound (Fig. 1e).

The impact of aP2 inhibition on atherosclerosis

To address whether inhibition of aP2 can alter the development of vascular lesions, we performed early and late intervention studies in the *Apoe^{-/-}* mouse model of atherosclerosis on a western diet. In the early intervention study, a western diet and the aP2 inhibitor BMS309403 were started simultaneously in 5-week-old mice (Supplementary Fig. 2a). In the late intervention paradigm, the aP2 inhibitor was administered after 8 weeks of a western diet (at 12 weeks of age), when significant atherosclerosis has developed (Fig. 2a). Analysis of the *en face* aorta demonstrated marked reductions in atherosclerotic lesion area in the aP2-inhibitor-treated group compared with the vehicle group in both the early (52.6%, Supplementary Fig. 2a, b) and late (51.0%, Fig. 2a, b) intervention studies. Staining of cross-sections of the proximal aorta with Oil Red O revealed fatty streak lesions (Fig. 2c and Supplementary Fig. 2c). These were almost exclusively macrophage-derived foam cells, as determined by immunohistochemical staining with MOMA-2 (Fig. 2d and Supplementary Fig. 2d). Macrophages were located predominantly on the luminal surface of the lesions. The extent of

atherosclerotic lesion area in the proximal aorta was significantly reduced in the aP2-inhibitor-treated group compared with vehicle-treated controls in both the early (Supplementary Fig. 2e) and late (Fig. 2e) intervention studies.

The aP2 inhibitor did not influence body weight, systemic glucose or lipid metabolism in *Apoe^{-/-}* mice (Supplementary Table 1). Examination of the distribution of cholesterol among the serum lipoprotein fractions by size-exclusion chromatography revealed similar lipoprotein profiles between the groups with a large peak in the very low density lipoprotein fractions and a reduced high density lipoprotein peak that was due to APOE-deficiency in both the early (Supplementary Fig. 2f) and late (Fig. 2f) intervention studies. No significant difference in glucose levels during glucose tolerance tests was observed between the vehicle and aP2 inhibitor groups (Supplementary Fig. 3a, b). These results are consistent with previous observations made in mice with genetic deficiency of aP2 in the *Apoe^{-/-}* background^{9,13}.

Cholesterol and inflammatory responses in macrophages

Macrophage foam cell formation has a critical role in the pathogenesis of atherosclerosis and is a process regulated by FABPs⁹. Transformation of THP-1 macrophage to foam cells was significantly reduced in the presence of aP2 inhibitor (25 μ M) (Fig. 3a). The aP2-inhibitor-treated THP-1 macrophages exhibited 44% reduction in

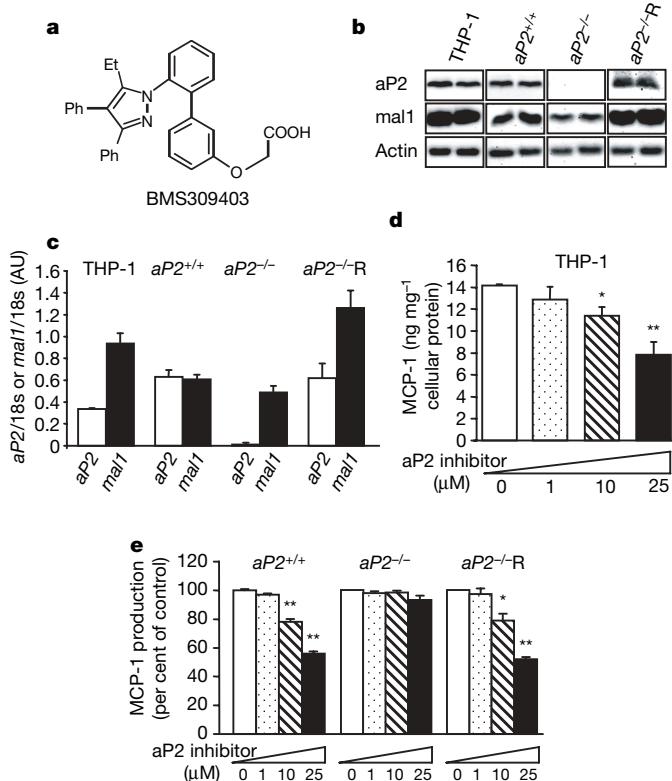


Figure 1 | Target-specific effects of aP2 inhibition on MCP-1 production in macrophages. **a**, Structure of the compound, BMS309403. **b**, Protein levels of aP2 and mal1 in human THP-1 macrophages and mouse macrophage cell lines, *aP2^{+/+}*, *aP2^{-/-}* and *aP2^{-/-R}*. **c**, aP2 and mal1 mRNA levels analysed by quantitative real-time PCR. **d**, MCP-1 production in human THP-1 macrophages treated with aP2 inhibitor at the indicated concentrations for 24 h. **e**, MCP-1 production in mouse cell lines treated with the aP2 inhibitor at the indicated concentrations for 24 h. Data are shown as the mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ compared with the control (each untreated cell line). AU, arbitrary units.

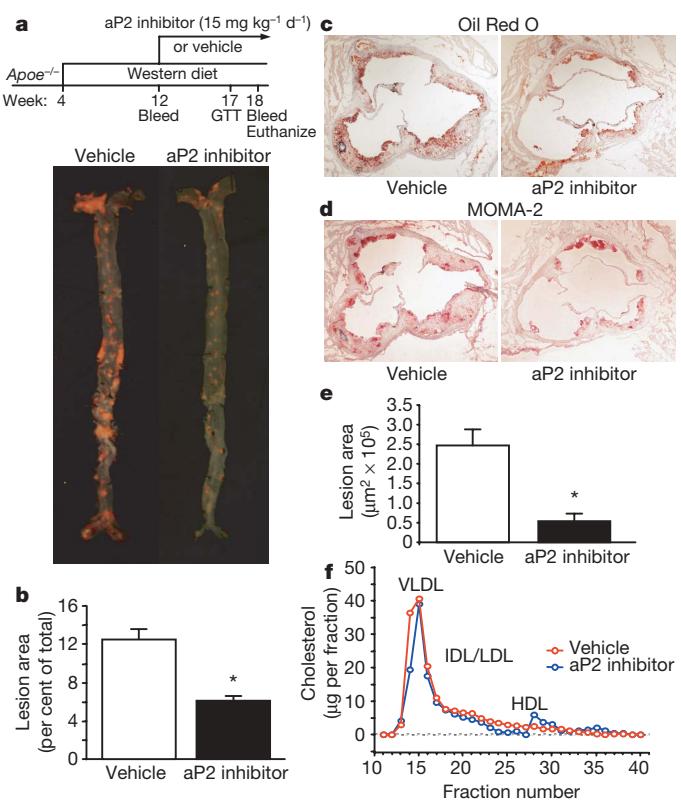


Figure 2 | Atherosclerosis in *Apoe^{-/-}* mice treated with the aP2 inhibitor. **a**, Experimental design of the late intervention study and *en face* aortas stained with Sudan IV. **b**, Quantitative analyses of the atherosclerotic lesion areas (per cent of total aorta surface area) in the vehicle ($n = 16$) and aP2 inhibitor ($n = 15$) groups. **c**, **d**, Oil Red O (**c**) and MOMA-2 (**d**) stainings of atherosclerotic lesions in the aortic root at the level of the aortic valves. Magnification, $\times 40$. **e**, Quantitative analyses of the proximal aorta atherosclerotic lesion areas in the vehicle ($n = 11$) and aP2 inhibitor ($n = 6$) groups. **f**, Lipoprotein profile in *Apoe^{-/-}* mice treated with vehicle (red) and aP2 inhibitor (blue) in the late intervention study. Data are presented as an average ($n = 3$) per cent distribution of total cholesterol for each group. Data are expressed as the mean \pm s.e.m. * $P < 0.01$. VLDL, very low density lipoprotein; IDL, intermediate density lipoprotein; LDL, low density lipoprotein; HDL, high density lipoprotein. GTT, glucose tolerance test.

cholesterol ester accumulation compared with vehicle-treated macrophages (Fig. 3b). In $aP2^{+/+}$ and $aP2^{-/-}R$ macrophages treated with the aP2 inhibitor, total cellular cholesterol ester content was significantly lower than in macrophages treated with vehicle (Fig. 3c). The aP2 inhibitor did not affect cholesterol ester content in the $aP2^{-/-}$ macrophages, again demonstrating the specificity of aP2 inhibitor action in this context.

To determine the potential mechanism for the reduction in cholesterol ester accumulation on inhibition of aP2, we examined APOA1-mediated cholesterol efflux in these cells. Cholesterol efflux from human THP-1 macrophages was significantly increased on treatment with the aP2 inhibitor (Fig. 3d). There was a significant increase in both mRNA and protein levels of the ATP-binding cassette A1 (ABCA1) protein, a critical mediator of cholesterol efflux in macrophages, in the aP2-inhibitor-treated THP-1 cells compared with vehicle-treated controls (Supplementary Fig. 4a, b). Consistent with earlier observations¹⁵, cholesterol efflux in $aP2^{-/-}$ macrophages was substantially higher than that of the $aP2^{+/+}$ cells (Fig. 3e) and was completely abrogated with reconstitution of aP2 expression. Similar to genetic deficiency, treatment with the aP2 inhibitor significantly increased cholesterol efflux in the $aP2^{+/+}$ and $aP2^{-/-}R$ macrophages but not in the $aP2^{-/-}$ cells. We also examined the

impact of aP2 inhibition on principal target molecules that regulate cellular cholesterol ester synthesis and hydrolysis in macrophages. There was a modest reduction in acyl-coenzyme A: cholesterol-acyl-transferase 1 (ACAT1), a key enzyme of cholesterol esterification, in the aP2-expressing macrophages (Fig. 3f) but no effect of aP2 inhibition on the expression of hormone-sensitive lipase, which acts as the neutral cholesterol esterase, in macrophages (Supplementary Fig. 5).

Macrophages participate in the pathogenesis of atherosclerosis not only through the formation of foam cells but also by the production of inflammatory mediators. Hence, we determined the impact of aP2 inhibition on several critical chemoattractant and inflammatory cytokines, including MCP-1, interleukin (IL)1 β , IL6 and tumour necrosis factor (TNF) in macrophages. Expression of these cytokines was significantly reduced in the aP2-expressing macrophages treated with the aP2 inhibitor compared with those treated with vehicle (Fig. 3g–j). No regulation was evident in $aP2^{-/-}$ cells on treatment with the inhibitor, demonstrating the target specificity of the aP2 inhibitor.

Inhibition of aP2 in adipocytes

The main site of aP2 expression is the adipocyte and although this site does not play a major part in atherosclerosis, it does significantly

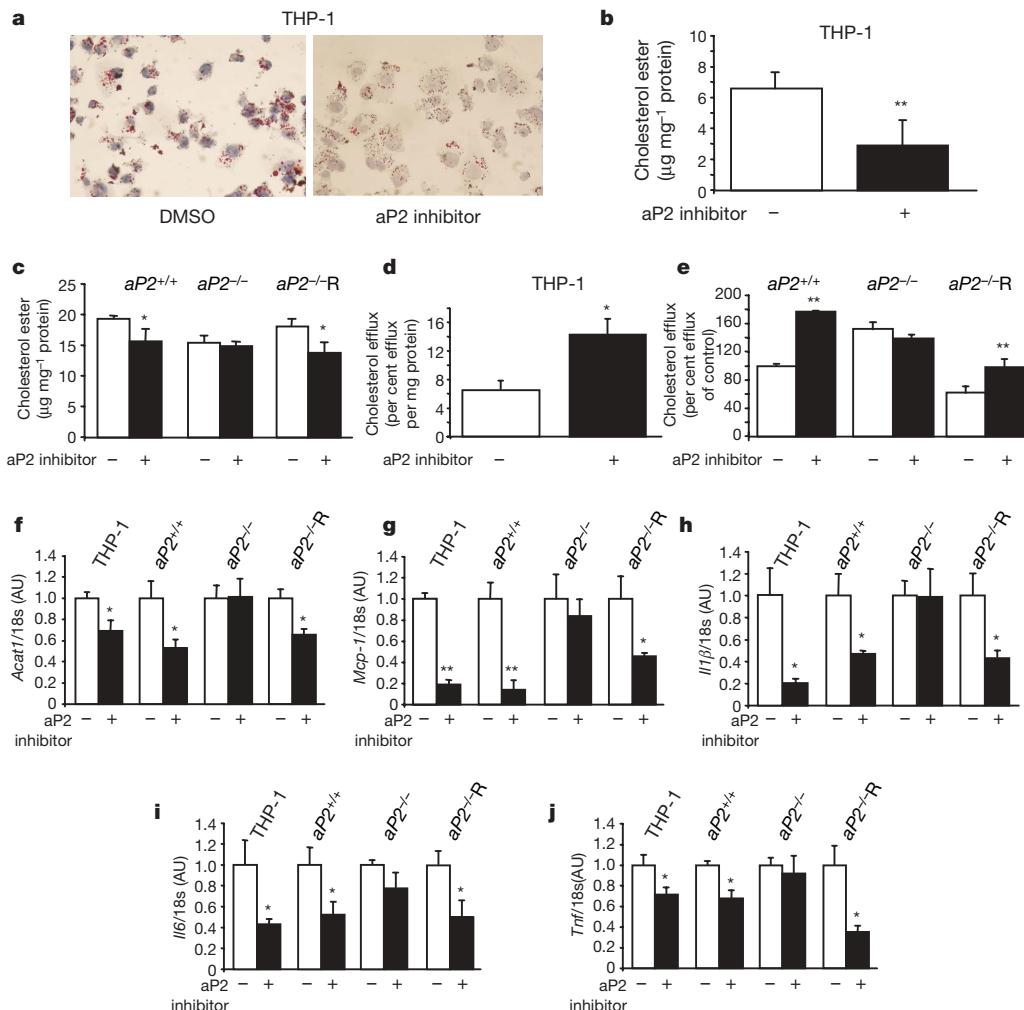


Figure 3 | Effects of aP2 inhibitor on lipid accumulation, cholesterol efflux and inflammatory responses in macrophages. **a**, Oil Red O staining of THP-1 macrophage foam cells loaded with acetylated low density lipoprotein ($50 \mu\text{g ml}^{-1}$) in the absence or presence of aP2 inhibitor ($25 \mu\text{M}$). Magnification, $\times 400$. **b**, **c**, Cholesterol ester levels normalized to cellular protein content in human THP-1 macrophages (**b**) and mouse macrophage cell lines, $aP2^{+/+}$, $aP2^{-/-}$ and $aP2^{-/-}R$ (**c**).

d, **e**, APOA1-specific cholesterol efflux in THP-1 macrophages (**d**) and mouse cell lines (**e**) in the absence or presence of aP2 inhibitor ($25 \mu\text{M}$). **f–j**, Expression of *Acat1* (**f**) and chemoattractant and inflammatory cytokines, *Mcp-1* (**g**), *Il1β* (**h**), *Il6* (**i**), and *Tnf* (**j**) in macrophages normalized to 18s rRNA levels. Data are normalized to untreated cells and expressed as the mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ compared with the control (each untreated cell line). DMSO, dimethyl sulphoxide.

contribute to systemic insulin resistance and type 2 diabetes^{6,7,9}. To begin to address the specific action of aP2 inhibition in adipocytes, we generated wild-type (*aP2*^{+/+} *mal1*^{+/+}) and FABP-deficient (*aP2*^{-/-} *mal1*^{-/-}) pre-adipocyte cell lines as well as FABP-deficient cells reconstituted with exogenous aP2 or with control empty vector. These cell lines fully differentiate into adipocytes and in all properties tested, behave in a similar way to commonly used 3T3-L1 or 3T3-F442A adipocytes (Fig. 4a). In genetic aP2-deficiency, the principal alteration observed in adipocytes thus far is a reduction in fatty acid transport⁸. Hence, we asked whether the chemical inhibition of aP2 could mimic this action in adipocytes and do so in an aP2-dependent fashion. Treatment with the aP2 inhibitor resulted in a dose-dependent decrease in fatty acid uptake in wild-type adipocytes (Fig. 4b). In contrast, there was no action of this inhibitor at any dose in FABP-deficient adipocytes. However, when aP2-deficient cells were reconstituted with aP2, they were rendered responsive to aP2 inhibition in a dose-dependent manner. Hence, the action of the synthetic aP2 inhibitor of regulating lipid transport in adipocytes was specific to its target, aP2.

Inhibition of aP2 in obese and diabetic mice

Having established the target specificity of BMS309403 in adipocytes, we administered the compound into a genetic model of obesity and insulin resistance, the leptin-deficient *ob/ob* (also known as *Lep*^{ob/ob}) mouse, and investigated insulin sensitivity and glucose metabolism. During the course of the 6-week treatment, there was no significant difference in body weight between animals receiving vehicle or aP2 inhibitor (Supplementary Fig. 6a). Similarly, per cent body fat, rates of oxygen consumption and carbon dioxide production, food intake, and physical activity were not different between the vehicle and aP2 inhibitor treatment groups (Supplementary Fig. 6b–f). In contrast, blood glucose levels in both the fed and fasted state were decreased after treatment with the aP2 inhibitor (Fig. 4c). Similar to genetic aP2-deficiency on the *ob/ob* background⁷, free fatty acid levels showed a trend towards an increase after treatment with the aP2 inhibitor (Supplementary Table 2, $P = 0.07$). The aP2 inhibitor decreased insulin and triglyceride levels and increased adiponectin concentration (Fig. 4d, e, and Supplementary Table 2), suggesting a potential increase in systemic insulin sensitivity. In fact, glucose tolerance tests revealed a significant improvement in glucose metabolism in the aP2-inhibitor-treated group (Fig. 4f). Similarly, insulin tolerance tests showed significantly increased insulin sensitivity in the *ob/ob* mice treated with the aP2 inhibitor (Fig. 4g). At the end of the treatment period, we analysed islet morphology in the pancreas. There was no difference in the pancreatic morphology such as the size, shape, and organization of the non-β-cell mantle between the vehicle- and aP2-inhibitor-treated *ob/ob* mice (Supplementary Fig. 7).

We also investigated the effect of aP2 inhibition in a diet-induced obesity model using both wild-type and FABP-deficient mice. The aP2-inhibitor-treated wild-type mice showed a significant decrease in glucose levels during glucose tolerance tests compared with vehicle-treated animals, but there was no change in glucose levels between the vehicle- and aP2-inhibitor-treated FABP-deficient mice on a high-fat diet (Supplementary Fig. 8). These results demonstrate that the insulin-sensitizing effects of the aP2 inhibitor *in vivo* are target-specific and effective in two independent models of obesity and insulin resistance.

Furthermore, we performed hyperinsulinaemic-euglycaemic clamp studies in *ob/ob* mice after 4 weeks of treatment. There was no significant difference in basal hepatic glucose production between the vehicle and aP2-inhibitor groups, but clamp hepatic glucose production was significantly suppressed in the aP2-inhibitor-treated *ob/ob* mice compared with vehicle-treated controls (Fig. 4h). Both whole-body glucose disposal and glucose infusion rates were also significantly increased after treatment with the aP2 inhibitor (Fig. 4i). These data demonstrate that the aP2 inhibitor improves whole-body insulin sensitivity through the suppression of hepatic glucose

production and enhancement of insulin-stimulated glucose disposal in peripheral tissues. To explore this further, we determined the rate of glucose uptake in gastrocnemius muscle and epididymal fat during the clamp procedure. In the aP2-inhibitor-treated mice, glucose uptake in muscle and adipose tissues was significantly increased compared with that in the vehicle-treated controls (Fig. 4j).

Effects of aP2 inhibition on adipose tissue in *ob/ob* mice

Adipocyte size in *ob/ob* mice treated with the vehicle or aP2 inhibitor was comparable, but macrophage infiltration in adipose tissue was more severe in the vehicle-treated group (Fig. 5a). Expression of two macrophage markers, *F4/80* (*Emr1*) and *Cd68*, was significantly

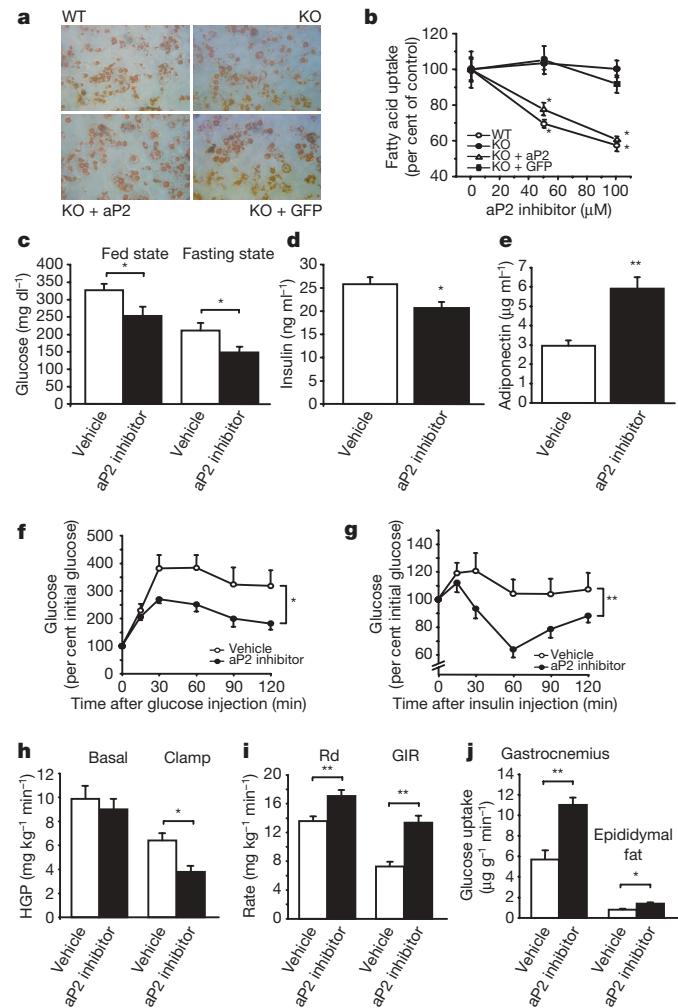


Figure 4 | Metabolic studies in aP2-inhibitor-treated adipocytes and *ob/ob* mice. **a**, Oil Red O staining of wild-type (WT), FABP-deficient (KO), FABP-deficient reconstituted with aP2 (KO + aP2), and FABP-deficient with vector (KO + GFP) adipocyte cell lines. **b**, Fatty acid uptake using ³H-stearate in adipocyte cell lines. **c**, Blood glucose levels in *ob/ob* mice treated with vehicle ($n = 6$) or aP2 inhibitor ($n = 6$) at the fed state after 2 weeks of treatment and at the fasting state after 6 weeks of treatment. **d, e**, Plasma levels of insulin (**d**) and adiponectin (**e**) in *ob/ob* mice treated with vehicle ($n = 6$) or aP2 inhibitor ($n = 6$) for 6 weeks. **f, g**, Glucose tolerance tests performed after 4 weeks of treatment in *ob/ob* mice with vehicle (open circle, $n = 6$) or aP2 inhibitor (closed circle, $n = 6$). **g**, Insulin tolerance tests performed after 5 weeks of treatment in *ob/ob* mice with vehicle (open circle, $n = 6$) or aP2 inhibitor (closed circle, $n = 6$). **h–j**, Hyperinsulinaemic-euglycaemic clamp studies performed in *ob/ob* mice treated with vehicle ($n = 7$) or aP2 inhibitor ($n = 9$) for 4 weeks. Basal and clamp hepatic glucose production (HGP) (**h**), glucose disposal rate (Rd) and glucose infusion rate (GIR) (**i**), and tissue glucose uptake in gastrocnemius muscle and epididymal fat (**j**). Data are shown as the mean \pm s.e.m.

* $P < 0.05$, ** $P < 0.01$.

reduced in the aP2-inhibitor-treated mice compared with vehicle-treated controls (Fig. 5b, c). Obesity leads to increased production of several chemoattractant and inflammatory cytokines, which have a critical role in obesity-associated inflammation and metabolic pathologies. Expression of *Mcp-1*, *Il1 β* , *Il6* and *Tnf* in adipose tissue was significantly reduced in *ob/ob* mice treated with the aP2 inhibitor compared with those treated with the vehicle (Fig. 5d–g).

Obesity-induced Jun N-terminal kinase (JNK) 1 activity is critical in the generation of inflammatory responses and inhibition of insulin action^{16,17}. To examine whether aP2 inhibition modifies the inflammatory profile and insulin action by this mechanism, we determined JNK1 activity in the adipose tissue of vehicle- and aP2-inhibitor-treated *ob/ob* mice. There was a significant attenuation (40%) of obesity-induced adipose tissue JNK1 activity in mice treated with aP2 inhibitor compared with vehicle-treated controls (Fig. 5h).

We next examined whether inhibition of aP2 and the alterations seen in inflammatory mediators and JNK activity in adipose tissue resulted in enhanced insulin action at this site. Insulin receptor signalling capacity was examined biochemically in intact mice following insulin administration. Insulin-stimulated tyrosine 1162/1163 phosphorylation of insulin receptor β subunit (IR β) or IRSR β subunit and serine 473 phosphorylation of AKT were significantly increased in the adipose tissue of aP2-inhibitor-treated *ob/ob* mice compared with that of vehicle-treated controls (Fig. 5i). These results demonstrate that aP2 inhibition reduced inflammation and increased insulin sensitivity in the adipose tissues of *ob/ob* mice.

Effects of aP2 inhibition on liver in *ob/ob* mice

In genetic aP2-deficiency, there is a striking molecular compensation through increased expression of mal1 in adipose tissue⁶. Owing to this compensation, the phenotype of aP2-deficiency is much milder than aP2-mal1 combined deficiency¹⁸. We show here that the compensatory increase in mal1 expression of adipose tissue in the genetic

absence of aP2 also occurs in the *ob/ob* background (Fig. 6a). Because the aP2-inhibitor-treated animals exhibit a significant protection against metabolic disease, we asked whether the compensatory increase in mal1 expression was present or not under these circumstances. Interestingly, after 6 weeks of the aP2-inhibitor treatment, there was no change in levels of aP2 or mal1 protein in the adipose tissue (Fig. 6a). This is a critical observation contributing to the efficacy of chemical inhibition of aP2 action in adult animals. For example, in genetic aP2-deficiency, there is no protection against fatty liver disease but a profound protection is seen in aP2-mal1 combined deficiency^{18,19}. In the aP2-inhibitor-treated *ob/ob* mice, fatty infiltration of the liver was attenuated (Fig. 6b) with a significant reduction in total liver triglyceride content (Fig. 6c). This reduction in fatty liver disease on aP2 inhibition was associated with diminished expression of key lipogenic enzymes in liver, including the stearoyl-CoA desaturase 1 (*Scd1*), fatty acid synthase (*Fasn*), and acetyl-CoA carboxylase 1 (*Acaca*) (Fig. 6d–f). This phenotype is reminiscent of aP2-mal1 combined deficiency rather than isolated aP2-deficiency^{18,19}.

In a similar way to adipose tissue, total JNK1 activity in the liver tissue of aP2-inhibitor-treated *ob/ob* mice was significantly reduced (43%) compared with that of vehicle-treated control mice (Fig. 6g). Suppression of fatty liver infiltration and inflammatory responses in aP2-inhibitor-treated *ob/ob* mice also resulted in enhanced insulin action in the liver. Insulin-stimulated tyrosine 1162/1163 phosphorylation of IR β and serine 473 phosphorylation of AKT were significantly increased in the liver tissue of aP2-inhibitor-treated *ob/ob* mice compared with vehicle-treated controls (Fig. 6h).

Discussion

A principal mechanistic core of obesity, type 2 diabetes and atherosclerosis resides at the interface of metabolic and inflammatory pathways¹. However, this mechanistic platform has not yet been exploited for the development of effective therapeutic strategies.

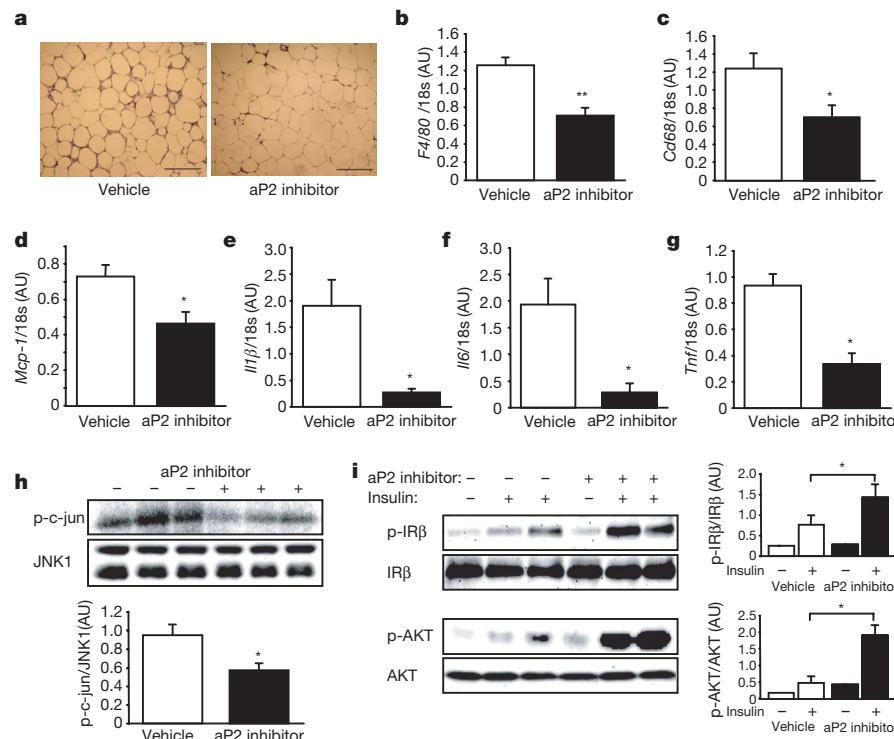


Figure 5 | Effects of aP2 inhibitor in adipose tissue of *ob/ob* mice.

a, Haematoxylin and eosin staining of the adipose tissue in *ob/ob* mice treated with vehicle or aP2 inhibitor. Scale bar, 200 μ m. **b–g**, Expression of *F4/80* (**b**), *Cd68* (**c**), *Mcp-1* (**d**), *Il1 β* (**e**), *Il6* (**f**), and *Tnf* (**g**) in the adipose tissue of *ob/ob* mice treated with vehicle ($n = 6$) or aP2 inhibitor ($n = 6$).

h, JNK1 activity in the adipose tissue of *ob/ob* mice. Quantification is shown in the graph below. **i**, Insulin-stimulated IR β tyrosine 1162/1163 and AKT serine 473 phosphorylation (p) in the adipose tissues of *ob/ob* mice. The graphs on the right of each blot show the quantification. Data are shown as the mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$.

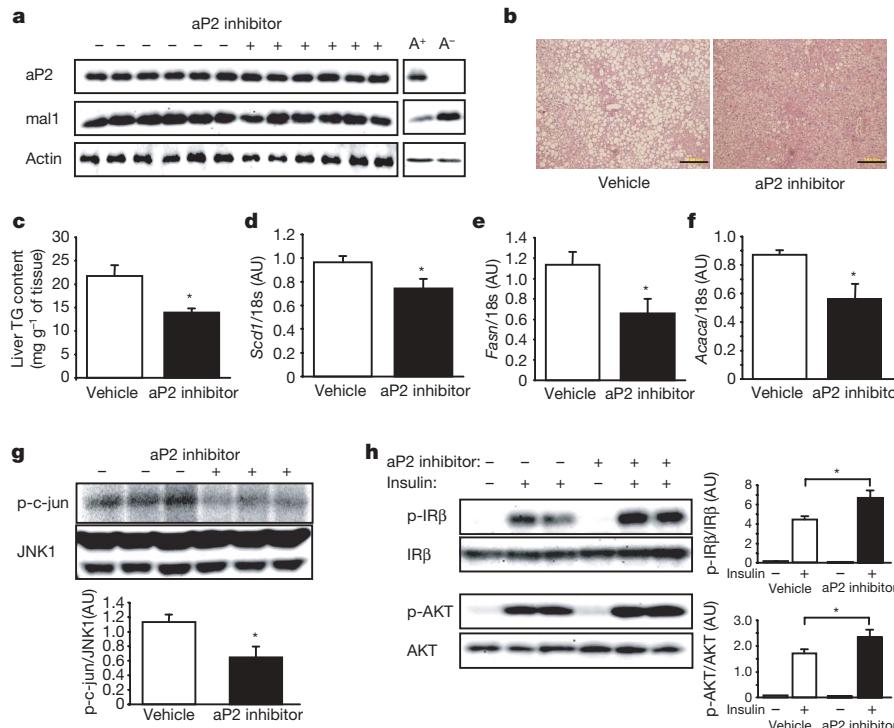


Figure 6 | Effects of aP2 inhibitor in liver of *ob/ob* mice. **a**, aP2 and mal1 protein in the adipose tissue of *ob/ob* mice treated with vehicle or aP2 inhibitor. For control, the adipose tissue of *ob/ob; aP2^{+/+}* (*A⁺*) and *ob/ob; aP2^{-/-}* (*A⁻*) mice was used. **b**, Haematoxylin and eosin staining of the liver of *ob/ob* mice treated with vehicle or aP2 inhibitor. Scale bar, 200 μ m. **c–f**, Triglyceride (TG) content (**c**) and mRNA expression of *Scd1* (**d**), *Fasn* (**e**), and *Acaca* (**f**) in the liver of *ob/ob* mice treated with vehicle ($n = 6$) or aP2

The locus for the adipocyte/macrophage FABP aP2 is critical in the regulation and dysregulation of metabolic and inflammatory responses as they relate to metabolic diseases^{6–9,13,15}. In addition to cell-autonomous effects in macrophages and adipocytes, aP2 also acts to coordinate the functional interactions between these two critical cell types in adipose tissue. Genetic deletion of the *aP2* gene in mice has demonstrated a strong role for this molecule in several chronic metabolic diseases, most notably, atherosclerosis and type 2 diabetes, and has raised the possibility of using aP2 as a potential drug target. Here, we have provided a critical proof of principle in mice that aP2 could be successfully targeted by an orally active, small-molecule inhibitor to generate a profile reminiscent of genetic deficiency *in vitro* and *in vivo*.

There are several indications that FABPs might be involved in metabolic homeostasis in a similar fashion in humans. First, the expression and regulation patterns of aP2 are very similar in both adipocytes and macrophages between mice and humans^{2,6,9}. Expression of aP2 is highly regulated on macrophage activation and, interestingly, suppressed by a statin *in vitro*²⁰. In both mouse and human macrophages, aP2 expression modulates inflammatory responses, foam cell formation and cholesterol efflux^{9,15}. Atherosclerotic lesions express high levels of aP2 in both mice and humans^{9,13,21}. Finally, aP2 expression is increased in obesity²². Hence, it is possible that aP2 function may be similar in humans as well. In fact, in a recent study, we produced genetic support for this concept in humans²³. A rare genetic variant was identified at the promoter region of the human *aP2* orthologue *FABP4*, coinciding with the binding site for C/EBP. This particular mutation alters C/EBP binding and significantly reduces the transcriptional activity of the human *aP2* promoter and the expression level of aP2 in the tissues of the carriers. In a large population sampling, individuals with the *aP2* variant had lower triglyceride levels, exhibited reduced cardiovascular disease risk and were protected from

inhibitor ($n = 6$). **g**, JNK1 activity in the liver of *ob/ob* mice treated with vehicle or aP2 inhibitor. The graph below the blot shows quantification. **h**, Insulin-stimulated IR β tyrosine 1162/1163 and AKT serine 473 phosphorylation in the liver tissues of *ob/ob* mice treated with vehicle or aP2 inhibitor. The graphs demonstrate the quantification of phosphorylation of each molecule. Data are shown as the mean \pm s.e.m. * $P < 0.05$.

obesity-induced type 2 diabetes. This study offers a critical insight and indicates that the metabolic function of aP2 in humans may be similar to that observed in mouse models. It is therefore possible that chemical inhibition of aP2 in humans might also show beneficial effects against diabetes and cardiovascular disease.

METHODS SUMMARY

The synthetic agent BMS309403 is a selective, high-affinity inhibitor of aP2. Information on synthesis and chemical properties has recently been reported¹⁴. Human monocytic leukaemia THP-1 cells were obtained from ATCC. Immortalized *aP2^{+/+}* and *aP2^{-/-}* mouse macrophage cell lines were generated in our laboratory as described^{9,15}. Wild-type (*aP2^{+/+} mal1^{+/+}*) and FABP-deficient (*aP2^{-/-} mal1^{-/-}*) pre-adipocytes were developed using a previously described protocol²⁴. The FABP-deficient pre-adipocytes were reconstituted by lentivirus with exogenous aP2 or with control empty vector including green fluorescent protein. All *Apoe^{-/-}* and *ob/ob* mice were from Jackson Laboratory, and *aP2^{-/-} mal1^{-/-}* mice were generated as previously described¹⁸. All mice have the C57BL/6J genetic background. The aP2 inhibitor was administered by oral gavage. Quantification of atherosclerotic lesions was performed as previously described²⁵. Hyperinsulinaemic-euglycaemic clamps were performed by modification of a described procedure²⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 November 2006; accepted 12 April 2007.

Published online 6 June 2007.

1. Hotamisligil, G. S. Inflammation and metabolic disorders. *Nature* 444, 860–867 (2006).
2. Hertzel, A. V. & Bernlohr, D. A. The mammalian fatty acid-binding protein multigene family: molecular and genetic insights into function. *Trends Endocrinol. Metab.* 11, 175–180 (2000).
3. Hunt, C. R., Ro, J. H., Dobson, D. E., Min, H. Y. & Spiegelman, B. M. Adipocyte P2 gene: developmental expression and homology of 5'-flanking sequences among fat cell-specific genes. *Proc. Natl. Acad. Sci. USA* 83, 3786–3790 (1986).

4. Melki, S. A. & Abumrad, N. A. Expression of the adipocyte fatty acid-binding protein in streptozotocin-diabetes: effects of insulin deficiency and supplementation. *J. Lipid Res.* **34**, 1527–1534 (1993).
5. Distel, R. J., Robinson, G. S. & Spiegelman, B. M. Fatty acid regulation of gene expression. Transcriptional and post-transcriptional mechanisms. *J. Biol. Chem.* **267**, 5937–5941 (1992).
6. Hotamisligil, G. S. et al. Uncoupling of obesity from insulin resistance through a targeted mutation in aP2, the adipocyte fatty acid binding protein. *Science* **274**, 1377–1379 (1996).
7. Uysal, K. T., Scheja, L., Wiesbrock, S. M., Bonner-Weir, S. & Hotamisligil, G. S. Improved glucose and lipid metabolism in genetically obese mice lacking aP2. *Endocrinology* **141**, 3388–3396 (2000).
8. Scheja, L. et al. Altered insulin secretion associated with reduced lipolytic efficiency in aP2^{-/-} mice. *Diabetes* **48**, 1987–1994 (1999).
9. Makowski, L. et al. Lack of macrophage fatty-acid-binding protein aP2 protects mice deficient in apolipoprotein E against atherosclerosis. *Nature Med.* **7**, 699–705 (2001).
10. Kazemi, M. R., McDonald, C. M., Shigenaga, J. K., Grunfeld, C. & Feingold, K. R. Adipocyte fatty acid-binding protein expression and lipid accumulation are increased during activation of murine macrophages by toll-like receptor agonists. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1220–1224 (2005).
11. Fu, Y., Luo, N. & Lopes-Virella, M. F. Oxidized LDL induces the expression of ALBP/aP2 mRNA and protein in human THP-1 macrophages. *J. Lipid Res.* **41**, 2017–2023 (2000).
12. Pelton, P. D., Zhou, L., Demarest, K. T. & Burris, T. P. PPAR γ activation induces the expression of the adipocyte fatty acid binding protein gene in human monocytes. *Biochem. Biophys. Res. Commun.* **261**, 456–458 (1999).
13. Boord, J. B. et al. Adipocyte fatty acid-binding protein, aP2, alters late atherosclerotic lesion formation in severe hypercholesterolemia. *Arterioscler. Thromb. Vasc. Biol.* **22**, 1686–1691 (2002).
14. Sulsky, R. et al. Potent and selective biphenyl azole inhibitors of adipocyte fatty acid binding protein (aFABP). *Bioorg. Med. Chem. Lett.* (in the press).
15. Makowski, L., Brittingham, K. C., Reynolds, J. M., Suttles, J. & Hotamisligil, G. S. The fatty acid-binding protein, aP2, coordinates macrophage cholesterol trafficking and inflammatory activity. Macrophage expression of aP2 impacts peroxisome proliferator-activated receptor γ and I κ B kinase activities. *J. Biol. Chem.* **280**, 12888–12895 (2005).
16. Hirosumi, J. et al. A central role for JNK in obesity and insulin resistance. *Nature* **420**, 333–336 (2002).
17. Tuncman, G. et al. Functional *in vivo* interactions between JNK1 and JNK2 isoforms in obesity and insulin resistance. *Proc. Natl. Acad. Sci. USA* **103**, 10741–10746 (2006).
18. Maeda, K. et al. Adipocyte/macrophage fatty acid binding proteins control integrated metabolic responses in obesity and diabetes. *Cell Metab.* **1**, 107–119 (2005).
19. Cao, H. et al. Regulation of metabolic responses by adipocyte/macrophage fatty acid-binding proteins in leptin-deficient mice. *Diabetes* **55**, 1915–1922 (2006).
20. Llaverias, G. et al. Atorvastatin reduces CD68, FABP4, and HBP expression in oxLDL-treated human macrophages. *Biochem. Biophys. Res. Commun.* **318**, 265–274 (2004).
21. Fu, Y., Luo, N., Lopes-Virella, M. F. & Garvey, W. T. The adipocyte lipid binding protein (ALBP/aP2) gene facilitates foam cell formation in human THP-1 macrophages. *Atherosclerosis* **165**, 259–269 (2002).
22. Fisher, R. M. et al. Fatty acid binding protein expression in different adipose tissue depots from lean and obese individuals. *Diabetologia* **44**, 1268–1273 (2001).
23. Tuncman, G. et al. A genetic variant at the fatty acid-binding protein aP2 locus reduces the risk for hypertriglyceridemia, type 2 diabetes, and cardiovascular disease. *Proc. Natl. Acad. Sci. USA* **103**, 6970–6975 (2006).
24. Sethi, J. K. et al. Characterisation of receptor-specific TNF α functions in adipocyte cell lines lacking type 1 and 2 TNF receptors. *FEBS Lett.* **469**, 77–82 (2000).
25. Babaev, V. R., Patel, M. B., Semenkovich, C. F., Fazio, S. & Linton, M. F. Macrophage lipoprotein lipase promotes foam cell formation and atherosclerosis in low density lipoprotein receptor-deficient mice. *J. Biol. Chem.* **275**, 26293–26299 (2000).
26. Kim, J. K. et al. Redistribution of substrates to adipose tissue promotes obesity in mice with selective insulin resistance in muscle. *J. Clin. Invest.* **105**, 1791–1797 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by grants from the NIH and the American Diabetes Association. M.F. is supported by a JSPS Postdoctoral Fellowship for Research Abroad from the Japan Society for the Promotion of Science. G.T. is supported by a fellowship from the Iacobucci Foundation.

Author Contributions G.S.H. designed and supervised experiments and analysed data. M.F. designed and performed experiments and analysed data. G.T., C.Z.G., E.V. and K.K. performed experiments. L.M. and G.A. developed cell lines from mice. V.R.B., S.F. and M.F.L. analysed lipoprotein profiles and advised on experiments. R.S., J.A.R. and R.A.P developed the aP2 inhibitor, BMS309403. M.F. and G.S.H wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.S.H. (ghotamis@hsph.harvard.edu).

METHODS

Biochemical reagents. All biochemical reagents were purchased from Sigma-Aldrich (Saint Louis) unless indicated.

The compound of aP2 inhibitor BMS309403. A synthetic agent, BMS309403, was developed and provided by Bristol-Myers Squibb Pharmaceutical Research Institute. This agent is a selective, high affinity inhibitor of aP2 with the following chemical properties: 2-(2'-(5-ethyl-3,4-diphenyl-1H-pyrazol-1-yl)biphenyl-3-yloxy) acetic acid (BMS309403): m.p. 174–176 °C; ¹H-NMR (500 MHz, CDCl₃): δ 9.20 p.p.m. (br s, 1H), 7.65 (d, J = 7.1 Hz, 1H), 7.6–7.5 (m, 3H), 7.4 (m, 2H), 7.3–7.1 (m, 5H), 7.04 (d, J = 16.6 Hz, 2H), 6.90 (d, J = 7.7 Hz, 1H), 6.87 (dd, J = 2.2, 8.2 Hz, 1H), 6.66 (s, 1H), 4.36 (s, 2H), 2.05 (br s, 2H), 0.58 (t, J = 7.2 Hz); ¹³C-NMR (125 MHz, CDCl₃): δ 172.2, 157.36, 149.1, 145.1, 139.5, 139.0, 137.2, 133.6, 132.7, 130.3, 130.2, 129.7, 129.5, 128.6, 128.4, 128.1, 128.0, 127.5, 126.8, 122.0, 119.0, 114.9, 113.6, 64.6, 17.6, 13.0; IR (KBr): 1710 cm⁻¹; analysis (% calculated, % found for C₃₁H₂₆N₂O₃): C (78.46, 78.30), H (5.52, 5.51), N (5.90, 5.69). Additional information on synthesis and chemical properties has recently been reported¹⁴.

Cells. Human monocytic leukaemia THP-1 cells were obtained from ATCC and cultured in Gibco RPMI 1640 medium (Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (Hyclone), 50 U ml⁻¹ penicillin and 50 µg ml⁻¹ streptomycin (Invitrogen) at 37 °C in 5% CO₂. THP-1 monocytes were differentiated into macrophages with 100 nM phorbol 12-myristate 13-acetate for 24 h. Immortalized aP2^{+/+} and aP2^{-/-} mouse macrophage cell lines were generated in our laboratory by a modification of a described procedure (refs 9, 27). Reconstitution of aP2 expression into aP2^{-/-} macrophages to produce the aP2^{-/-}R cells was performed as described¹⁵. The levels of aP2 protein, as assessed by western blot, were similar in the aP2^{-/-}R cell line as compared with the aP2^{+/+} macrophage line. Human THP-1 macrophages and mouse macrophage cell lines, aP2^{+/+}, aP2^{-/-} and aP2^{-/-}R, were incubated in RPMI 1640 supplemented with heat-inactivated 10% FBS or 5% lipoprotein-deficient serum (Biomedical Technologies) in the absence or presence of aP2 inhibitor dissolved in dimethyl sulphoxide (DMSO) at the indicated concentrations. The incubation periods varied according to the experimental protocol. Each experiment was done in at least triplicate.

We generated aP2^{+/+} mal1^{+/+} (WT) and aP2^{-/-} mal1^{-/-} (KO) pre-adipocytes from mouse models using a previously described protocol²⁴. The KO pre-adipocytes were also reconstituted by lentivirus with exogenous aP2 (KO+aP2) or with control empty vector including green fluorescent protein (KO+GFP). These cell lines were maintained and propagated in Dulbecco's Modified Eagle's Media (Invitrogen) with 10% cosmic calf serum (Hyclone), 50 U ml⁻¹ penicillin and 50 µg ml⁻¹ streptomycin at 37 °C in 10% CO₂. Differentiation was then initiated (day 0) by incubation in induction medium (1 µM dexamethasone, 0.5 mM isobutylmethylxanthine, 1 µM rosiglitazone and 5 µg ml⁻¹ insulin). Following a 4-day induction period (two 48-h incubations), the medium was changed to a post-induction medium (1 µM rosiglitazone and 5 µg ml⁻¹ insulin) for an additional 2 days. Thereafter, the medium was replaced with a medium supplemented only with 0.5 µg ml⁻¹ of insulin.

Mice. Animal care and experimental procedures were performed with approval from animal care committees of Harvard University. Male Apoe^{-/-} mice in the C57BL/6J background (Jackson Laboratory) were kept on a 12-h light cycle and were fed a high-cholesterol atherogenic western diet (D12079B: 21% fat, 0.21% cholesterol; Research Diets) *ad libitum*, beginning at 4–5 weeks of age. The mice were treated by oral gavage with vehicle including 10% 1-methyl-2-pyrrolidone and 5% cremophor EL with ethanol in 100 µl of water or 15 mg kg⁻¹ d⁻¹ of the aP2 inhibitor BMS309403 dissolved in the vehicle for 6 weeks starting at 5- (Supplementary Fig. 2a, early intervention) or 12-weeks of age (Fig. 2a, late intervention). Before and after treatment, blood samples were collected from animals that had fasted for 6 h, and the glucose tolerance test was performed by intraperitoneal glucose injection (1.5 g kg⁻¹) on conscious mice after an overnight (15 h) fast at the fifth week of treatment.

Male ob/ob mice (Jackson Laboratory) were treated by oral gavage with 40 mg kg⁻¹ d⁻¹ of the aP2 inhibitor or vehicle for 6 weeks beginning at 8-weeks-old. Before and after treatment, blood samples were collected from mice that had fasted for 6 h. Blood glucose levels at the fed state were also examined after 2 weeks of treatment. After 4 weeks of treatment, glucose tolerance test was performed by intraperitoneal glucose injection (0.5 g kg⁻¹) after an overnight fast. After 5 weeks of treatment, insulin tolerance test was performed by intraperitoneal insulin injection (2 IU kg⁻¹) after a 6 h fast.

Mice deficient in both aP2 and mal1 were generated as previously described¹⁸. These mice were backcrossed >12 generations into the C57BL/6J genetic background. Both wild-type and aP2^{-/-} mal1^{-/-} mice were placed on a high-fat diet (D12492: 60% kcal% fat; Research Diets), beginning at 4 weeks of age. In these mice, the aP2 inhibitor (40 mg kg⁻¹ d⁻¹) or vehicle was administered by oral

gavage for 4 weeks starting at 20 weeks of age. Glucose tolerance test was performed by intraperitoneal glucose injection (1.5 g kg⁻¹) after an overnight fast. **Assessment of atherosclerosis and immunohistochemistry.** Perfusion fixation, preparation of aortas and quantification of atherosclerotic lesions were performed as previously described²⁵. *In face* pinned-out aortas were stained with Sudan IV. The heart with aorta was embedded in OCT and snap-frozen in liquid nitrogen. Lesions in the proximal aorta from serial 7-µm-thick cryosections were stained with Oil Red O and counterstained with haematoxylin. The images of the aortas were captured with a digital colour camera (DP70; Olympus) mounted on a microscope (VistaVision). Quantitative analysis of lipid-stained lesions was performed using ImageJ software. Macrophages in the arterial lesions were examined by immunohistochemistry using 7-µm cryosections of the proximal aorta fixed in acetone at 4 °C. The sections were immersed in phosphate-buffered saline (PBS) and incubated overnight at 4 °C with monoclonal rat anti-MOMA-2 (Accurate Chemical & Scientific). The sections were treated with rabbit biotinylated antibodies to rat IgG for 1 h at room temperature followed by incubation with avidin-biotin complex labelled with alkaline phosphatase (Vectastain ABC-AP) and visualized with red alkaline phosphatase substrate (Vector Laboratories). Slides were counterstained with haematoxylin.

Plasma measurements. Blood glucose concentration was determined with 3-µl whole blood using Ascensia BREEZE blood glucose meter (Bayer Company). Plasma insulin was measured with a commercially available ultra sensitive ELISA kit (Crystal chemicals). Total cholesterol, free fatty acids, glycerol and triglycerides were determined with colorimetric assay systems (Wako chemicals and Sigma-Aldrich) adapted for microtitre plate assay. Plasma adiponectin was measured with an ELISA kit (Lincore Research). Particle size distribution of the lipoproteins was determined by fast-performance liquid chromatography (FPLC), using pooled samples of plasma as previously described²⁵. In brief, plasma from mice was subjected to FPLC analysis using a Superose 6 column (Pharmacia Biotech) on an HPLC system model 600 (Waters Chromatography). A 100-µl aliquot of plasma was injected onto the column and separated with a buffer containing 0.15 M NaCl, 0.01 M Na₂HPO₄, 0.1 mM EDTA (pH 7.5), at a flow rate of 0.5 ml min⁻¹. Forty 0.5-ml fractions were collected, and tubes 11–40 were analysed for cholesterol. Fractions 14–17 contain VLDL and chylomicra; fractions 18–24 contain LDL and IDL; fractions 25–29 contain HDL; and fractions 30–40 contain non-lipoprotein-associated proteins.

Examination of macrophage MCP-1 production. Supernatants from macrophages incubated in RPMI 1640 supplemented with 5% LPDS in the presence of aP2 inhibitor (0–25 µM) for 24 h were evaluated for MCP-1 production with ELISA (OptEIA; BD Biosciences). MCP-1 levels were normalized to cellular protein content in THP-1 macrophages. In the macrophage cell lines, aP2^{+/+}, aP2^{-/-} and aP2^{-/-}R, data were also normalized to those in the absence of aP2 inhibitor.

Intracellular lipid analysis in macrophages. THP-1 macrophages were incubated for 24 h in RPMI 1640 supplemented with 5% LPDS and 50 µg ml⁻¹ of acLDL (Biomedical Technologies), followed by RPMI 1640 supplementation with 10% FBS with the vehicle or aP2 inhibitor (25 µM) dissolved in DMSO. After 3 days, the macrophage foam cells were fixed in 10% formalin for 90 min, washed thoroughly with PBS, and incubated with a working solution of Oil Red O for 3 h. Intracellular lipids in macrophages were extracted with hexane/isopropanol (3:2) after 4 days incubation with the vehicle or aP2 inhibitor (25 µM), evaporated and dissolved in isopropanol containing 10% Triton X-100 for preparation of a sample solution. Free cholesterol and total cholesterol were determined by commercial assay systems (Wako chemicals). Cholesterol ester was estimated by subtracting free cholesterol from total cholesterol. Data are normalized to cellular protein content.

Cholesterol efflux assays in macrophages. APOA1-specific cholesterol efflux to the medium was determined as described¹⁵. Macrophages were incubated for 24 h in RPMI 1640 supplemented with 5% LPDS, 1 µCi ml⁻¹ of ³H-cholesterol (PerkinElmer Life Sciences) and 50 µg ml⁻¹ of acLDL in the absence or presence of aP2 inhibitor (25 µM) for 24 h. To equilibrate cholesterol pools, cells were washed twice with PBS and incubated for 24 h in RPMI 1640 containing 0.2% low endotoxin fatty-acid-free bovine serum albumin with no lipoproteins in the absence or presence of aP2 inhibitor (25 µM). Cells were washed again with PBS and incubated in RPMI containing 0.2% BSA in the absence or presence of APOA1 (20 µg ml⁻¹; Calbiochem) for 4 h. Radioactivity of culture supernatants and cellular lysate was measured separately by liquid scintillation. Results are normalized to cellular protein content and expressed as the percentage of ³H-cholesterol in the medium divided by the total ³H-cholesterol in the cells and medium per milligram of protein. In the mouse macrophage cell lines, data are also normalized to those in the untreated aP2^{+/+} macrophages.

Fatty acid uptake in adipocytes. Differentiated adipocytes were incubated in DMEM supplemented with 2% FBS in the absence or presence of aP2 inhibitor at the indicated concentrations for 24 h. Medium was changed to DMEM

containing 2% BSA in the absence or presence of aP2 inhibitor. Cells were then incubated with ^3H -stearate (PerkinElmer Life Sciences) at 37 °C for 30 min. Radioactivity of cellular lysate was measured by liquid scintillation. Results are normalized to cellular protein content and expressed as the percentage of those in the absence of aP2 inhibitor in adipocytes.

Metabolic studies. Total body fat was assessed by dual energy X-ray absorptiometry (DEXA; PIXImus). For additional metabolic measurements, mice were placed in an indirect open circuit calorimeter (Oxymax System). Oxygen and carbon dioxide concentrations by volume were monitored at the inlet and outlet parts of a partially sealed chamber, through which a known flow of ambient air was forcibly ventilated. The concentration difference measured between the parts was used to compute oxygen consumption (VO_2) and carbon dioxide production (VCO_2). The consumption and production information were presented in units of $\text{ml kg}^{-1} \text{h}^{-1}$ and normalized to 25 °C and 760 mm Hg. Food intake was investigated by using the Oxymax Feed Scale Device (Columbus Instruments). The physical activity of the mice was monitored with OPTO-M3 Activity Application Device (Columbus Instruments). The movements (other than scratching, grooming, digging and so on) of each animal were determined by infrared beams in x , y and z axes.

Hyperinsulinaemic–euglycaemic clamp studies. Hyperinsulinaemic–euglycaemic clamps were performed in *ob/ob* mice after 4 weeks of treatment by a modification of a described procedure²⁶. Four days before experiments, *ob/ob* mice were anaesthetized with an intraperitoneal injection of ketamine (80 mg kg^{-1}) and xylazine (40 mg kg^{-1}), and the right jugular vein was catheterized with a PE-10 polyethylene tube (inside and outside diameters, 0.28 mm and 0.61 mm, respectively; Becton Dickinson) filled with heparin solution (100 U ml^{-1}) (United States Pharmacopeia). After an overnight fast, HPLC purified ^3H -glucose (0.05 $\mu\text{Ci min}^{-1}$; Perkin Elmer) was infused during the 2-h basal period, and blood samples were collected at the end to estimate the rate of basal hepatic glucose production. After the basal period, a 120-min hyperinsulinaemic–euglycaemic clamp was conducted with a primed-continuous infusion of human insulin (Humulin R; Eli Lilly) at a rate of 12.5 mU $\text{kg}^{-1} \text{min}^{-1}$. Blood samples were collected at 20-min intervals for the immediate measurement of plasma glucose concentration, and 25% glucose was infused at variable rates to maintain plasma glucose at basal concentrations. Insulin-stimulated whole-body glucose disposal was estimated with a continuous infusion of ^3H -glucose throughout the clamps (0.1 $\mu\text{Ci min}^{-1}$). All infusions were performed using microdialysis pumps (CMA/Microdialysis). To estimate insulin-stimulated glucose uptake in individual tissues, $2\text{-}^{14}\text{C}$ -deoxyglucose ($2\text{-}^{14}\text{C}$ -DG; Perkin Elmer) was administered as a bolus (10 μCi) 75 min after the start of clamps. Because 2-deoxyglucose is a glucose analogue that is phosphorylated but not metabolized, insulin-stimulated glucose uptake in individual tissues can be estimated by determining the tissue content of $2\text{-deoxyglucose-6-phosphate}$. Blood samples were collected at 80, 85, 90, 100, 110, and 120 min after the start of clamps for the determination of plasma ^3H -glucose, $^3\text{H}_2\text{O}$, and $2\text{-}^{14}\text{C}$ -DG concentrations. At the end of clamps, animals were euthanized. Within 5 min, gastrocnemius muscles from both hindlimbs and epididymal adipose tissue were harvested. Each tissue was frozen immediately using liquid N₂ and stored at -80 °C until further analysis.

For the determination of plasma ^3H -glucose and $2\text{-}^{14}\text{C}$ -DG concentrations, plasma was deproteinized with ZnSO₄ and Ba(OH)₂, dried to remove $^3\text{H}_2\text{O}$, resuspended in water and counted in scintillation fluid (Ecoscint H) on channels for ^3H and ^{14}C . The plasma concentration of $^3\text{H}_2\text{O}$ was determined by the difference between ^3H counts without and with drying. For the determination of tissue $2\text{-}^{14}\text{C}$ -DG-6-phosphate content, tissue samples were homogenized, and the supernatants were subjected to an ion-exchange column to separate $2\text{-}^{14}\text{C}$ -DG-6-phosphate from $2\text{-}^{14}\text{C}$ -DG.

Rates of basal hepatic glucose production and insulin-stimulated whole-body glucose uptake were determined as the ratio of the ^3H -glucose infusion rate to the specific activity of plasma glucose at the end of the basal period and during the final 30 min of clamps, respectively. Hepatic glucose production during the hyperinsulinaemic–euglycaemic clamps was determined by subtracting the glucose infusion rate from the whole-body glucose uptake. Glucose uptake in individual tissues was calculated from the plasma $2\text{-}^{14}\text{C}$ -DG profile, which was fitted with an exponential curve, and tissue $2\text{-}^{14}\text{C}$ -DG-6-phosphate content.

Portal vein insulin infusion and protein extraction from tissue. Following 6 h of food withdrawal, *ob/ob* mice were anaesthetized with an intraperitoneal injection of tribromoethanol (250 mg kg^{-1}), and insulin (2 IU kg^{-1}) or phosphate buffered saline (PBS) was injected into mice through the portal vein. Three minutes after infusion, tissues were removed and frozen in liquid nitrogen and kept at -80 °C until processing. For protein extraction, tissues were placed in a cold lysis buffer containing 50 mM Tris-HCl (pH 7.0), 2 mM EGTA, 5 mM EDTA, 30 mM NaF, 10 mM Na₃VO₄, 10 mM Na₄P₂O₇, 40 mM β -glycerophosphate, 0.5% NP-40 and 1% protease inhibitor cocktail. After

homogenization on ice, the tissue lysates were centrifuged, and the supernatants were used for western blot analysis or JNK kinase assay.

Western blot analysis. Total protein content of the samples was assessed by microplate protein assay (Bio-Rad), and equal amounts of protein per sample and known molecular weight markers were subjected to SDS-polyacrylamide gel electrophoresis. Proteins were electrophoretically transferred onto PVDF membranes (Whatman) and incubated for 1 h at room temperature with blocking solution (2% BSA or 5% nonfat milk in TBS buffer containing 0.1% Tween 20). The blocked membranes were incubated with anti-aP2, anti-mali, anti-ABCA1 (Novus Biologicals), anti-IR β -pTyr 1162/1163 (Calbiochem), anti-IR β (Santa Cruz Biotechnology), anti-AKT-pSer 473 (Santa Cruz Biotechnology), anti-AKT (Santa Cruz Biotechnology), anti-JNK1 (Santa Cruz Biotechnology), anti-tubulin (Santa Cruz Biotechnology), and anti- β -actin (Cell Signaling Technology) for overnight at 4 °C, and washed three times with TBS buffer containing 0.1% Tween 20. The membranes were incubated with the secondary antibody conjugated with horseradish peroxidase (Amersham Biosciences) for 1 h at room temperature and washed. Immunodetection analyses were accomplished using the enhanced chemiluminescence kit (Roche Diagnostics).

c-Jun N-terminal kinase (JNK) kinase assay. Tissue lysates containing 500–1,000 μg of protein were mixed with 1 μg of JNK1 antibody (Santa Cruz Biotechnology) and 20 μl of protein A sepharose beads. The mixture was agitated at 4 °C overnight, pelleted by centrifugation and washed twice with lysis buffer followed by 3 additional washes with JNK kinase assay buffer containing 25 mM Hepes (pH 7.4), 20 mM MgCl₂, 20 mM β -glycerophosphate, 0.5 mM EGTA, 0.5 mM NaF, 0.5 mM Na₃VO₄ and 1 mM PMSF for equilibration. After washing with kinase buffer, the beads were incubated in 25 μl kinase buffer containing 10 μCi ^{32}P γ ATP (PerkinElmer Life Sciences), 10 μM ATP and 5 μg of c-jun fusion protein at 30 °C for 20 min. The reaction was terminated by addition of Laemmli buffer.

Lipid content in liver. Total lipids in liver were extracted by Bligh-Dyer method (ref. 28). Triglyceride content was determined by the colorimetric method using a commercial kit (Sigma-Aldrich).

Quantitative real-time PCR analysis. Total RNA was isolated using Trizol reagent (Invitrogen). For reverse transcription, 0.5–1.0 μg of the total RNA was converted to first strand complementary DNA in 20 μl reactions using a cDNA synthesis kit (Bio-Rad). Quantitative real-time PCR analysis was performed using SYBR Green in a real-time PCR machine (iCycler; Bio-Rad). The thermal cycling program was 4 min at 95 °C for enzyme activation and 50 cycles of denaturation for 15 s at 95 °C, 30 s annealing at 58 °C and 30 s extension at 72 °C. Primers used in the present study are listed in Supplementary Table 3. To normalize expression data, 18s rRNA was used as an internal control gene. For some experiments, another housekeeping gene, GAPDH, was used, and the experiments produced identical results to those obtained when 18s rRNA was used as a reference.

Statistical analysis. Experimental results were shown as the mean \pm s.e.m. After analysis of normality using the Shapiro-Wilk W test, the mean values for biochemical data from each group were compared by Student's *t*-test. Data on vascular lesions, which were not normally distributed, were analysed with a non-parametric Mann-Whitney test. Comparisons between time points were analysed using repeated-measures analysis of variance, ANOVA. All statistical tests with $P < 0.05$ were considered significant.

- 27. Blasi, E. et al. Selective immortalization of murine macrophages from fresh bone marrow by a *raf/myc* recombinant murine retrovirus. *Nature* **318**, 667–670 (1985).
- 28. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).

ARTICLES

Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers

Richard S. Maser^{1*}, Bhudipa Choudhury^{4*}, Peter J. Campbell^{4*}, Bin Feng^{2*}, Kwok-Kin Wong¹, Alexei Protopopov², Jennifer O'Neil³, Alejandro Gutierrez^{3,5}, Elena Ivanova², Ilana Perna², Eric Lin⁶, Vidya Mani¹, Shan Jiang¹, Kate McNamara¹, Sara Zaghoul¹, Sarah Edkins⁴, Claire Stevens⁴, Cameron Brennan⁷, Eric S. Martin¹, Ruprecht Wiedemeyer¹, Omar Kabbarah¹, Cristina Nogueira¹, Gavin Histen⁸, Jon Aster⁸, Marc Mansour¹¹, Veronique Duke¹¹, Letizia Foroni¹¹, Adele K. Fielding¹¹, Anthony H. Goldstone¹², Jacob M. Rowe¹³, Yaoqi A. Wang^{1,2}, A. Thomas Look³, Michael R. Stratton⁴, Lynda Chin^{1,2,9}, P. Andrew Futreal⁴ & Ronald A. DePinho^{1,2,10}

Highly rearranged and mutated cancer genomes present major challenges in the identification of pathogenetic events driving the neoplastic transformation process. Here we engineered lymphoma-prone mice with chromosomal instability to assess the usefulness of mouse models in cancer gene discovery and the extent of cross-species overlap in cancer-associated copy number aberrations. Along with targeted re-sequencing, our comparative oncogenomic studies identified *FBXW7* and *PTEN* to be commonly deleted both in murine lymphomas and in human T-cell acute lymphoblastic leukaemia/lymphoma (T-ALL). The murine cancers acquire widespread recurrent amplifications and deletions targeting loci syntenic to those not only in human T-ALL but also in diverse human haematopoietic, mesenchymal and epithelial tumours. These results indicate that murine and human tumours experience common biological processes driven by orthologous genetic events in their malignant evolution. The highly concordant nature of genomic events encourages the use of genetically unstable murine cancer models in the discovery of biological driver events in the human oncogenome.

An increasing compendium of tumour genes underpins the development of effective, targeted therapeutic agents for cancer. This model now influences all aspects of cancer research, motivating intensive efforts in cancer gene discovery. Array-based comparative genome hybridization (array-CGH) has uncovered highly re-arranged human cancer genomes harbouring many recurrent copy number alterations (CNAs)¹. Targeted re-sequencing of human cancers has uncovered pathogenetic mutations in genes such as *BRAF* in melanoma² and *NOTCH1* in T-cell acute lymphoblastic leukaemias/lymphomas (T-ALL)³. Nonetheless, identifying true cancer genes, that is, differentiating 'causal' genomic events from bystander 'genomic noise', remains challenging.

A growing body of evidence, including cross-species preservation of oncogene transcriptional signatures⁴ and synteny of tumour-associated CNAs^{5–9}, indicates that mouse models of human cancer can serve as biologically relevant filters for complex human oncogenomes. More recent high-resolution genomics-based studies have proven that such cross-species comparisons enable the efficient identification of new oncogenes in human cancers^{10,11}, and re-sequencing of key genes in murine cancer models reveals conservation of somatic signature mutational events¹². However, a limitation of mouse cancer models for comparative oncogenomics is the relatively benign cytogenetic profiles of most genetically engineered murine cancers^{8,10,11,13}. On the other

hand, murine cancer genomes can be destabilized experimentally, to simulate the rampant chromosomal instability of human cancers, including those with telomere dysfunction, impaired DNA damage checkpoints such as Atm, dominant *Trp53* mutant alleles, and defective DNA repair^{14–21}. These 'instability models' of cancer, coupled with high-resolution genomic technologies, hold the potential to provide highly informative systems for comparative oncogenomics.

In this study, we generated a murine lymphoma model that combines the genome-destabilizing impact of Atm deficiency and telomere dysfunction to effect T lymphomagenesis in a *Trp53*-dependent manner. We applied multiple genome analysis technologies to identify cancer-associated alterations shared between our mouse model and human cancers. These systematic cross-species comparisons uncovered many syntenic CNAs, not only in lymphomas of the T-cell lineage, but also in other human tumour types, emphasizing the common genetic and biological events in mouse models of cancer. Re-sequencing of candidate genes in syntenic CNAs readily identified *FBXW7* and *PTEN* as commonly mutated genes in human T-ALL.

Results

Generation of murine T-cell lymphomas with highly complex genomes. Nearly all *Atm*^{−/−} mice develop thymic lymphomas²² with moderate levels of aneuploidy and recurrent chromosomal rearrangements²³. In

¹Department of Medical Oncology, ²Center for Applied Cancer Science of the Belfer Institute for Innovative Cancer Science, ³Department of Pediatric Oncology Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁴Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. ⁵Division of Hematology, Children's Hospital, Boston, Massachusetts 02115, USA. ⁶Agilent Technologies, Palo Alto, California 94304, USA. ⁷Department of Neurosurgery, Memorial Sloan Kettering Cancer Center, New York, New York 10021, USA. ⁸Department of Pathology, ⁹Department of Dermatology, ¹⁰Department of Genetics and Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹¹Royal Free and University College Medical School, London NW3 2PF, UK. ¹²University College London Hospitals, London NW1 2BU, UK. ¹³Rambam Medical Center and Technion, Haifa 31096, Israel.

*These authors contributed equally to this work.

Terc^{-/-} *Atm*^{-/-} mice, telomere dysfunction suppresses lymphoma development, which may relate in part to Trp53 activation^{20,21} and/or altered chromosomal dynamics²⁴. To address this directly, we intercrossed *Terc*, *Atm* and *Trp53* mutant mice to generate various genotypic combinations from this ‘triple’-mutant colony (hereafter designated as ‘TKO’ for all genotypes) and found that telomere dysfunctional (that is, third (G₃) and fourth (G₄) generation *Terc*^{-/-}; ref. 21) *Trp53*^{+/-} or *Trp53*^{-/-} TKO mice developed lymphoma with shorter latency and higher penetrance relative to *Trp53*^{+/-} TKO animals (Fig. 1a). Moreover, lymphomas from *Trp53*^{+/-} TKO mice showed reduction to homozygosity in 14/15 specimens examined (Fig. 1b), indicating strong genetic pressure to inactivate Trp53 during lymphomagenesis. These TKO tumours resembled those from conventional *Atm*^{-/-} mice with effacement of thymic architecture by CD4⁺/CD8⁺ (less commonly CD4⁻/CD8⁻ or mixed single/double positive) lymphoma cells (Fig. 1c, data not shown). Together, these observations strongly suggest that an *Atm*-independent *Trp53*-dependent telomere checkpoint constrains lymphoma development.

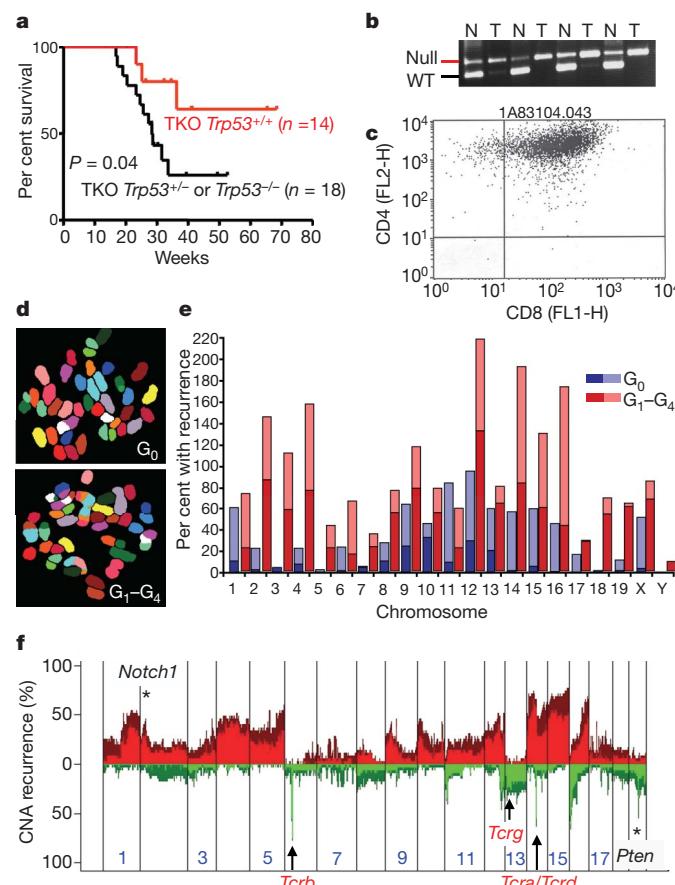


Figure 1 | The TKO model. **a**, Kaplan–Meier curve of thymic lymphoma-free survival for G₃–G₄ TKO mice with varying *Trp53* dosage. **b**, *Trp53* loss of heterozygosity assessed by PCR. N, normal; T, tumour. **c**, Representative FACS profile of a TKO tumour with CD4 and CD8 cell surface marker antibodies. **d**, Representative SKY images from metaphase spreads from G₀ (upper panel) and G₁–G₄ (lower panel) lymphomas. **e**, Quantification of cytogenetic aberrations (recurrences) detected by SKY in G₀ (blue) and G₁–G₄ (red) thymic lymphomas. Darker portion of column indicates proportion of non-reciprocal translocations; lighter portion indicates dicentric/robertsonian-like rearrangements. **f**, Recurrence plot of CNAs defined by array-CGH for 35 TKO lymphomas. The x axis shows the physical location of each chromosome. The percentage of tumours harbouring gains (dark red, $\log_2 \geq 0.3$), amplifications (bright red, $\log_2 \geq 0.6$), losses (green, $\log_2 \leq -0.3$), and deletions (dark green, $\log_2 \leq -0.6$) for each locus is depicted. Locations of physiologically relevant CNAs at *Tcrb*, *Tcra/Tcrd* and *Tcrg* (arrows), and *Notch1* and *Pten* (asterisks) are indicated.

We undertook spectral karyotype (SKY) analyses to quantify chromosomal rearrangements in nine telomere-deficient (G₁–G₄ *Terc*^{-/-}) and 9 telomere-intact (G₀ *Terc*^{+/-} or *Terc*^{+/-}) TKO lymphomas (Fig. 1d; Supplementary Fig. 1; Supplementary Table 1). Relative to G₀ tumours, G₁–G₄ TKO lymphomas showed a greater frequency of chromosome structural aberrations (0.34 versus 0.09 per chromosome, respectively, $P < 0.0001$, *t*-test), including multicentric chromosomes, non-reciprocal translocations, and p-p, p-q and q-q chromosome arm fusions involving homologous and/or non-homologous chromosomes. Several chromosomes (specifically, 2, 6, 8, 14, 15, 16, 17 and 19) were involved in significantly more dicentric and robertsonian-like rearrangements in G₁–G₄ relative to G₀ TKO tumours ($P < 0.05$; *t*-test; Fig. 1e), suggesting that these events provide adaptive mechanisms to tolerate telomere dysfunction and/or have causal roles in lymphoma development.

TKO tumours acquire genomic lesions syntenic to those in human T-ALL. In the mouse, cancer-associated chromosomal rearrangements of the types described above are by-products of breakage–fusion–bridge cycles associated with telomere dysfunction and defective Trp53 function¹⁴, and similar events can generate amplifications or deletions at the translocation site in human cells²⁵. These processes, together with biological selection, drive emergence of clonal CNAs targeting cancer-relevant genes in the mouse^{6,15}. We therefore compiled high-resolution genome-wide array-CGH profiles of 35 TKO tumours (Supplementary Table 1) and 26 human T-ALL cell lines and tumours (Supplementary Table 2) for comparison. TKO profiles revealed marked genome complexity with all chromosomes exhibiting regional and focal recurrent CNAs (Fig. 1f), some of which were observed in >40% of samples (for example, amplicons targeting distinct regions on mouse chromosomes 1, 2, 3, 4, 5, 9, 10, 12, 14, 15, 16, and 17; and deletions on 6, 11, 12, 13, 14, 16 and 19), corresponding well with rearrangements detected by SKY.

In addition to recurrent physiological deletions of T-cell receptor (*Tcr*) loci (Fig. 1f, arrows; Supplementary Fig. 2) as expected for T-cell derived tumours, we observed pathogenetically relevant recurrent genomic events in this instability model, such as a high-amplitude genomic event on chromosome 2. By SKY, this is a recurrent non-reciprocal translocation involving chromosome 2 (band A3) with different partner chromosomes (Supplementary Fig. 3). By array-CGH, the CNAs in 4 independent TKO tumours shared a common boundary close to the 3' region of the *Notch1* gene, with 2 additional tumours harbouring *Notch1* focal amplifications (Supplementary Fig. 3; data not shown). NOTCH1 activation by carboxy-terminal structure alteration and point mutations is a signature event of human T-ALL^{3,26,27}. Although the murine genomic rearrangements differ from *NOTCH1* translocations in human T-ALL²⁷, their common shared boundary involving *Notch1* suggested potential relevance to the TKO tumours. Thus, we performed *Notch1* re-sequencing in several TKO lymphomas lacking *Notch1* genomic rearrangement and uncovered truncating insertion/deletion mutations and non-conservative amino acid substitutions in the *Notch1* PEST and heterodimerization domains, as well as broader deletions within exon 34 encoding the PEST domain (Supplementary Fig. 4a; Supplementary Table 1). This mutation spectrum is analogous to that observed in human T-ALL, because the PEST and heterodimerization domains are two hot spots of *NOTCH1* mutation (Supplementary Fig. 4a, see below)³. Biochemically, these mutations promote activation of mouse Notch1 as shown by higher levels of both full-length and the active cleaved form (V1744)²⁸ of Notch1 proteins (Supplementary Fig. 4b). Transcriptome profiling further showed upregulation of several Notch1 transcriptional targets including *Ptcr*, *Hes1*, *Dtx1* and *Cd3e*, which correlated well with messenger RNA levels of *Notch1*²⁶ (Supplementary Fig. 4c).

The observation of physiological deletion of *Tcr* loci and a human-like pattern of *Notch1* genomic and mutational events in TKO tumours prompted the assessment of the extent to which other CNAs emerging in TKO lymphomas mirrored those seen in human T-ALL. To this end, we first defined the minimal common regions

(MCRs, regions overlapping in more than one tumour as defined in Methods) in the TKO genome using an established algorithm with criteria of CNA width ≤ 10 Mb (megabases) and amplitude > 0.75 (\log_2 scale), yielding 160 MCRs with average sizes of 2.12 Mb (0.15–9.82 Mb) and 2.33 Mb (0.77–9.6 Mb) for amplifications and deletions, respectively (Supplementary Table 3). This frequency of genomic alterations is comparable to most human cancer genomes (for example, see below) and significantly above the typical 20 to 40 events detected in most genetically engineered ‘genome-stable’ murine tumour models^{8,10,11,29}. Next, by orthologue mapping of genes resident within the MCRs of CNAs, we compared the TKO MCRs with similarly defined MCRs in human T-ALL to identify syntenic overlap. Here, 18 of the 160 MCRs (11%) were found to overlap across species (Table 1); the extent to which overlap occurred was significantly greater than that expected by chance alone ($P = 0.001$ and 0.004 for deletions and amplifications, respectively). Moreover, these syntenic MCRs included several genes already implicated in T-ALL biology, such as *Crebbp*, *Ikaros* (also known as *Ikzf1*) and *Abi1* (refs 30–32). These data support the relevance of murine models to human cancer biology and their potential use for human cancer gene discovery through comparative oncogenomics.

Frequent *Fbxw7* inactivation in T-ALL. A few TKO tumours with minimal Notch1 expression exhibited elevated *Notch4* or *Jagged1* (Notch ligand) mRNA levels (data not shown), prompting a detailed examination of the genomic and expression status of components in the Notch pathway. Indeed, a syntenic MCR encompassing the *Fbxw7* gene (MCR no. 18, Table 1; Fig. 2a) was observed across species. Although extremely focal in humans (Fig. 2a), the syntenic overlap made it unlikely that such deletion events represented copy number polymorphisms. *FBXW7* re-sequencing in a cohort of human T-ALL clinical specimens ($n = 38$) and cell lines ($n = 23$) (Supplementary Tables 2a, c and 4) revealed that *FBXW7* was mutated or deleted in 11/23 of the cell lines (48%) and 11/38 of the clinical samples (29%), making this one of the most commonly mutated genes in human T-ALL (Supplementary Tables 4 and 5). Consistent with reduced expression of *Fbxw7* relative to non-neoplastic thymus in 19 of the 24 TKO lymphomas (Fig. 2b), these *FBXW7* mutations in human T-ALL were predominantly mis-sense mutations, and particularly clustered in evolutionarily conserved arginine residues of the third and fourth WD40 domains involved in target binding³³ (Fig. 2c).

Moreover, re-sequencing of *FBXW7* in matched normal bone marrows from several patients in complete remission showed that the two most common mutations observed (R465, R479) were acquired somatically (data not shown). Additional mutations are likely to be somatic because the majority have been reported as somatic mutations in other tumour types (<http://www.sanger.ac.uk/genetics/CGP/cosmic/>) and none were present in public single nucleotide polymorphism databases. Finally, 19 of 21 mutations were heterozygous, consistent with reports that *Fbxw7* acts as a haplo-insufficient tumour suppressor gene³⁴.

FBXW7 is a component of the E3 ubiquitin ligase responsible for binding the NOTCH1 PEST domain, leading to ubiquitination and proteasome degradation³⁵. PEST domain mutations in human T-ALL prolong the half-life of intracellular NOTCH1, indicating that *FBXW7* loss may cause similar phenotypic effects. To address this, we characterized the cell lines and clinical samples for *NOTCH1* mutations (Supplementary Tables 2a, 2c, 4 and 5). Interestingly, there was no association between known functional mutations of *NOTCH1* (N-terminal (HD-N) and C-terminal (HD-C) heterodimerization domains and PEST domains) and *FBXW7* mutations ($P = 0.16$). However, among samples with *NOTCH1* mutations, *FBXW7* mutations were found less frequently in samples with a mutated PEST domain (4/19; 21%) than samples with mutations of only the HD-N or HD-C domain (13/20; 65%; $P = 0.009$). This is consistent with the hypothesis that mutations of *FBXW7* and the PEST domain of *NOTCH1* target the same degradation pathway, and little selective advantage accrues to the majority of tumours from mutating both components. On the other hand, the lack of *NOTCH1* and *FBXW7* mutation mutual exclusivity suggests that *FBXW7* affects other pathways in addition to NOTCH (see discussion).

Frequent *Pten* inactivation in mouse and human T-cell malignancy. Focal deletion on chromosome 19, centring on *Pten*, was among the most common genomic events in TKO lymphomas (Table 1; Fig. 1f). Array-CGH, coupled with real-time genomic PCR, documented homozygous deletions of *Pten* in 15/35 (43%) TKO lymphomas (Fig. 3; Supplementary Fig. 5). *PTEN* is a well-known tumour suppressor³⁶ and the inactivation of the homologue in the murine thymus is known to generate T-cell tumours³⁷. Correspondingly, array-CGH confirmed that 4 out of 26 human T-ALL samples had *PTEN* locus rearrangements. Additionally, re-sequencing of the 61 T-ALL cell lines and clinical specimens (Supplementary Table 2)

Table 1 | Synteny mapping between murine TKO lymphoma and human T-ALL MCRs

MCR no.	Cytoband	Start	End	Mouse TKO			Cancer genes or candidates	Chr	Start	Human T-ALL		
				Size (bp)	Peak ratio	Rec				End	Size (bp)	Peak ratio
Amplification												
1	4E2	153,362,787	154,677,539	1,314,752	0.88	13	<i>Dvl1</i> ; <i>Ccnl2</i> ; <i>Aurkaip1</i>	1	1,286,939	153,633,5	249,396	1.11
2	10A3	18,124,375	22,105,516	3,981,141	1.91	11	<i>Myb</i> ; <i>Ahi1</i>	6	135,471,648	135,829,074	357,426	1.07
3	16C4	91,250,715	97,408,345	6,157,630	1.38	21	<i>Runx1</i> ; <i>Ets2</i> ; <i>Tmprss2</i> ; <i>Ripk4</i> ; <i>Erg</i>	21	40,837,575	422,856,61	1,448,086	0.95
4	5G2	136,128,574	138,413,308	2,284,734	0.87	14	<i>Gnb2</i> ; <i>Perq1</i>	7	99,901,102	99,949,527	48,425	1.09
5	4A1	5,601,642	13,568,807	7,967,165	1.00	11	<i>Tox</i>	8	59,880,732	60,101,149	220,417	0.82
6	2B	29,315,580	31,992,174	2,676,594	1.78	7	<i>Set</i> ; <i>Fnbp1</i> ; <i>Abi1</i> ; <i>Nup214</i>	9	130,710,910	131,134,550	423,640	2.06
Deletion												
7	11B3-B4	68,759,068	72,041,187	3,282,119	-0.93	4	<i>Trp53</i> ; <i>Bcl6b</i>	17	6,494,426	7,767,821	1,273,395	-0.76
8	3H4	155,474,073	158,861,389	3,387,316	-0.75	3	<i>Negr1</i>	1	71,919,083	72,444,137	525,054	-0.92
9	15B3.1	33,212,025	41,060,793	7,848,768	-0.93	2	<i>Baal</i> ; <i>Fzd6</i>	8	104,310,865	104,499,581	188,716	-0.93
10	16A1	3,264,231	10,275,117	7,010,886	-0.97	21	<i>Crebbp</i> ; <i>Ciita</i>	16	3,195,168	11,549,999	8,354,832	-1.09
11	19C3-D2	46,457,272	56,116,765	9,659,493	-0.77	8	<i>Mxi1</i>	10	111,672,720	112,043,485	370,765	-0.90
12	4E2	150,778,332	154,677,539	3,899,207	-0.83	2	<i>Hes3</i> ; <i>Rpl22</i> ; <i>Chd5</i>	1	5,983,967	6,318,619.5	334,652	-0.85
13	11A1	8,844,892	12,372,703	3,527,811	-3.73	14	<i>Ikaros</i>	7	49,539,939	50,229,252.5	689,313	-0.75
14	12F2	111,667,310	115,272,402	3,605,092	-1.43	9	<i>Ptpn2</i>	7	156,125,925	158,194,699	2,068,774	-0.84
15	6B1	41,191,601	41,690,238	498,637	-5.48	28	<i>Tcrb</i>	7	141,785,426	142,078,458	293,032	-3.07
16	19A	11,295,986	15,610,191	4,314,205	-0.77	4	<i>Gnaq</i>	9	77,572,992	77,916,022	343,030	-0.76
17	19C1	31,573,449	32,118,682	545,233	-4.48	13	<i>Pten</i>	10	89,594,719	90,035,234	440,515	-3.30
18	3E3-F1	79,297,034	87,003,791	7,706,757	-0.93	2	<i>Fbxw7</i>	4	153,078,068	154,979,435	1,901,367	-1.74

Each murine TKO MCR with synteny overlap with an MCR in the human T-ALL data set is listed—separated by amplification and deletion—along with its chromosomal location (Cytoband/Chr) and base number (Start and End, in base pairs). The minimal size of each MCR is indicated in base pairs (bp). Peak ratio refers to the maximal \log_2 array-CGH ratio for each MCR. Rec refers to the number of tumours in which the MCR was defined.

uncovered inactivating *PTEN* mutations in 9 cases, none of which was found in public SNP databases. Each is predicted to cause protein truncation and is observed in other cancers³⁶ but shows no clear correlation with *NOTCH1* mutations (Supplementary Tables 4 and 5). *PTEN* mutations were more frequent in cell lines (7/23; 30.4%) than in clinical specimens (2/38; 5.2%) (Supplementary Table 4). Because these clinical specimens were newly diagnosed cases, whereas the cell lines were mostly established from relapses, this difference in mutation frequency may suggest that *PTEN* inactivation is a later event associated with progression, among other possibilities.

Northern and western blot analyses and transcriptome profiling of the TKO and human T-ALL samples revealed a broader collection of tumours with low to undetectable *PTEN* expression (Fig. 3b, data not shown) with elevated phospho-AKT. Moreover, AKT activation may have been driven by focal Akt1 amplification and Tsc1 loss in two TKO samples (Fig. 3c; data not shown). Lastly, the biological significance of Pten status in TKO lymphoma is supported by their sensitivity to Akt inhibition in a Pten-dependent manner (Supplementary Fig. 6) in response to triciribine, a drug known to block Akt phosphorylation and inhibit cells dependent on the Akt pathway³⁸.

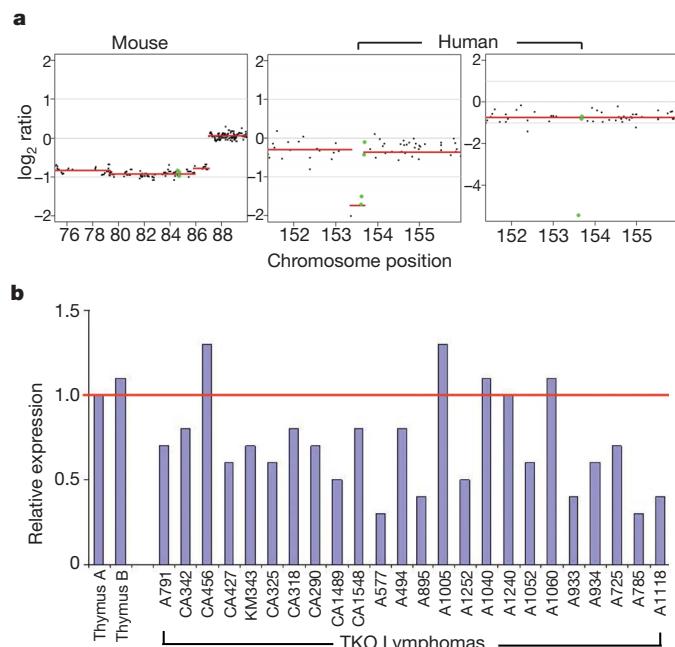


Figure 2 | *FBXW7* alterations are common in human T-ALL and conserved in the murine TKO tumours. **a**, log₂ ratio of array-CGH plots showing conserved deletion of *FBXW7* in both mouse TKO and human T-ALL cell lines; *FBXW7* is indicated in green. The y axis shows log₂ of copy number ratio (normal, log₂ = 0); amplifications are above and deletions are below this axis; x axis, chromosome position, in Mbp. **b**, Relative expression level (by real-time quantitative PCR) of mouse *Fbxw7* mRNA in murine TKO tumours, with expression in normal thymus (thymus A) set at 1; an independent normal thymus sample (thymus B) was run for comparison. **c**, Location of *FBXW7* mutations in a panel of human T-ALL patients and cell lines. Each marker represents an individual cell line/patient.

Broad comparison of the TKO genome with diverse human cancers. Although cross-species comparison showed numerous concordant lesions in cancers of T-cell origin, the fact that this instability model is driven by mechanisms of fundamental relevance (telomere dysfunction and Trp53 mutation) to many cancer types suggested potentially broader relevance to other human cancers. A case in point is the *PTEN* example above; *PTEN* is a tumour suppressor for multiple cancers³⁶. We therefore extended the cross-species comparative genomic analyses to 6 other human cancer types ($n = 421$) of haematopoietic, mesenchymal and epithelial origins, including multiple myeloma ($n = 67$)³⁹, glioblastoma ($n = 38$) (unpublished) and melanoma ($n = 123$) (unpublished), as well as pancreatic ($n = 30$) (unpublished), lung ($n = 63$)⁴⁰ and colon adenocarcinomas ($n = 74$) (unpublished).

Compared to similarly defined MCR lists (that is, MCR width ≤ 10 Mb; Fig. 4a) of each of these cancer types, we found that 102 (61 amplifications, 41 deletions) of 160 MCRs (64%) in the TKO genomes matched with at least 1 MCR in 1 human array-CGH data set, with strong statistical significance attesting to non-randomness of the overlap (Fig. 4a). Confidence in the relevance of these syntenic events was bolstered by the observation that more than half of these syntenic MCRs (38 of 61 amplifications or 62%; 22 of 41 deletions or

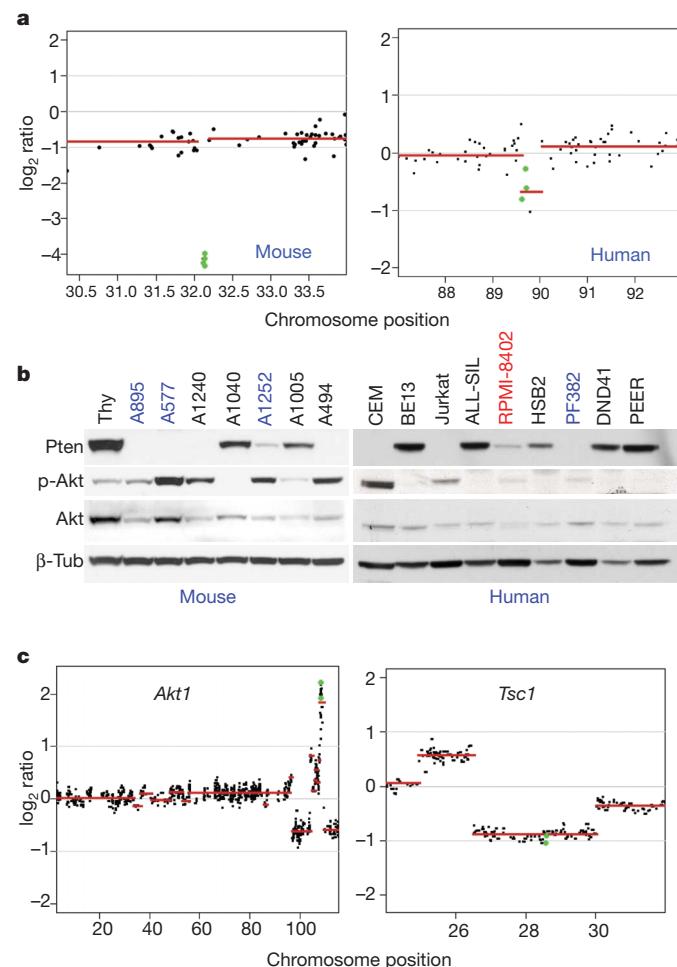


Figure 3 | Conservation of *PTEN* genetic alterations in human and mouse T-ALLs. **a**, Array-CGH plots showing conserved deletion of *PTEN* in both mouse TKO and human T-ALL cell lines; *PTEN* indicated in green. The y axis shows log₂ ratio of copy number (normal, log₂ = 0); amplifications are above and deletions are below the red axis. **b**, Western blotting for PTEN, phospho-Akt, Akt and tubulin (loading control) in a panel of murine TKO and human T-ALL cell lines. BE13 and PEER are synonymous lines. Samples in red harbour confirmed sequence mutations; samples in blue harbour aCGH-detected deletions. **c**, CNAs affecting other members of the *Pten*-Akt axis are shown as log₂ ratio plots for murine TKO tumours. The location of each gene (*Akt1/Tsc1*) is shown in green.

53%) overlapped with MCRs recurrent in two or more human tumour types (Fig. 4b). Moreover, a significant proportion of the TKO MCRs were evolutionarily conserved in human tumours of non-haematopoietic origin (Fig. 4c). Among the 61 amplifications with syntenic hits, 58 of them (95%) were observed in solid tumours, whereas the remaining 3 were uniquely found in myeloma (Fig. 4c). Similarly, 33 of 41 (80%) syntenic deletions were present in solid tumours (Fig. 4c). Among these are *Trp53*, deleted in 5 of 7 human cancer types, and *Myc* amplifications in 6 human cancers. This substantial overlap with diverse human cancers was unexpected.

Next, to provide an additional level of validation for these TKO genomic events, we determined whether these syntenic MCRs targeted known cancer genes. Mouse homologues were found for 237 of 363 genes listed on the Cancer Gene Census⁴¹, and of these, 24 were found within the 104 syntenic MCRs (Supplementary Table 6). These included 17 oncogenes in amplifications and 7 tumour suppressor genes in deletions. However, the majority of these syntenic MCRs do not contain known cancer genes, raising the strong possibility that resequencing that is focused on resident genes of syntenic MCRs may

provide a high-yield strategy to identify somatic mutations in human cancers, a thesis supported by the *FBXW7* and *PTEN* examples.

Discussion

We describe here the characterization of a mouse model of T-cell lymphoma engineered with telomere and checkpoint defects. In these unstable murine tumours, cytogenetic and copy number analyses reveal a level of genome complexity comparable to that of human solid tumours. The presence of rampant genome instability, and absence of an engineered dominantly acting oncogene, created a model system in which biological selection pressures drive acquired genomic events with potential relevance to human T-cell malignancy and/or cancer in general. In this model, murine cancers acquire widespread recurrent clonal CNAs targeting loci syntenic to alterations present in a human T-ALL as well as a large collection of diverse haematopoietic, mesenchymal and epithelial cancers. The relevance of these non-random syntenic events is supported on several levels, including the presence of syntenic MCRs in more than one human cancer type; the identification of MCRs harbouring known cancer genes; and the identification of somatic mutations in genes within conserved MCRs (*FBXW7* and *PTEN*).

Evolutionarily conserved genomic deletions targeting the *FBXW7* and *PTEN* loci motivated re-sequencing of them in human and murine T-ALL samples, revealing frequent mutations. *FBXW7* mutation has been described in about 5% of colorectal and ovarian cancers^{42–44}, although its involvement in T-ALL has not been reported previously. As a key component of the E3 ubiquitin ligase that ubiquitinates *NOTCH1*, loss or mutation of *FBXW7* is expected to be functionally equivalent to PEST domain mutation of *NOTCH1*, prolonging *NOTCH1* half-life³⁵. Indeed, a statistically significant anti-correlation is observed between presence of PEST domain mutations in *NOTCH1* and presence of *FBXW7* mutation in human T-ALL samples. However, the fact that several samples with *FBXW7* mutations have wild-type *NOTCH1* (5/22) strongly implies that other pathways may be affected by *FBXW7* mutations in some cancers. Indeed, *FBXW7* is known to promote degradation of other proteins implicated in lymphoma biology such as Aurora A (also known as *AURKA*), c-Jun (JUN), Cyclin E (*CCNE1*) and C-myc (*MYC*)^{34,35}, the latter of which is also a transcriptional target of NOTCH signalling⁴⁵.

In summary, the compelling synteny between TKO tumours and several human cancers of diverse origins demonstrates that engineered chromosomal instability in the mouse can engender genomic alterations similar to those observed in complex human oncogenomes, providing evidence that murine and human tumours experience common biological processes driven by the orthologous genetic events. These results provide support for the use of such murine models as a guidepost to focus and prioritize re-sequencing efforts in human cancer because cross-species synteny serves as a measure of validation by virtue of the evolutionary conservation and use of different genetic mechanisms (that is, mutation and copy number) involved. Furthermore, the mouse can facilitate the differentiation between somatic CNAs and copy number polymorphisms owing to the fragmentation of syntenic regions in the mouse and human genomes. We conclude that genetically unstable mouse cancer models represent a valuable resource for mining complex human cancer genomes.

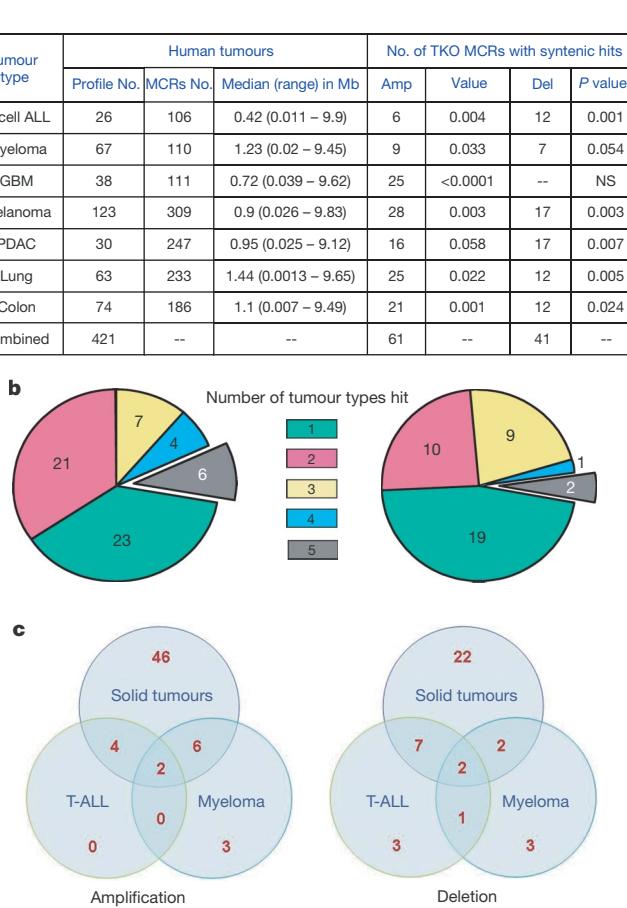


Figure 4 | Substantial overlap between genomic alterations of murine TKO lymphomas and human tumours of diverse origins. **a**, Characteristics of MCRs from human array-CGH profiles from the indicated tumour types are listed on the left portion of the panel. PDAC, pancreatic adenocarcinoma. The number of TKO MCRs (Amp, amplifications; Del, deletions) with syntenic overlap with the corresponding human CGH data set is indicated on the right side of the panel (P value is based on 10,000 permutations). **b**, Pie-chart showing numbers of TKO MCRs (indicated within each segment) with syntenic overlap identified in one or multiple human tumour types (indicated by different colours of the segments); left panel, amplifications; right panel, deletions. **c**, Venn diagram representation of overlap between murine TKO MCRs and MCRs from human cancers of T-ALL, multiple myeloma, or solid tumours (encompassing glioblastoma, melanoma, and pancreatic, lung and colon adenocarcinoma).

METHODS SUMMARY

Terc Atm- and *Terc Trp53*-deficient mice described previously^{21,46} were interbred, maintained in pathogen-free facilities and followed for lymphoma development. Resultant TKO lymphomas were harvested from moribund animals and metaphase preparations of primary cultures were used for SKY profiling. DNA and RNA were extracted from the murine TKO tumours as well as from human cancer cell lines and tumours for performing array-CGH and transcriptome profiling, respectively. MCRs of 10 Mb or smaller were defined by described algorithms^{39,47} for murine lymphomas and each of the 6 different human cancer types. Syntenic overlap was determined based on orthologue mapping of MCR resident genes and statistic significance of the overlap calculated by permutation. Known cancer genes were as defined by the Cancer Gene Census⁴¹ (<http://www.sanger.ac.uk/>)

genetics/CGP/Census). Gene mutation status was established by denaturing high-performance liquid chromatography as previously published⁴⁸, and by bidirectional sequencing. All eight array-CGH data sets used in this study are available on the GEO website under the accession number GSE7615.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 March; accepted 30 April 2007.

Published online 21 May 2007.

- Pinkel, D. & Albertson, D. G. Array comparative genomic hybridization and its applications in cancer. *Nature Genet.* **37** (Suppl) S11–S17 (2005).
- Davies, H. et al. Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
- Weng, A. P. et al. Activating mutations of *NOTCH1* in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).
- Sweet-Cordero, A. et al. An oncogenic *KRAS2* expression signature identified by cross-species gene-expression analysis. *Nature Genet.* **37**, 48–55 (2005).
- Hodgson, G. et al. Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. *Nature Genet.* **29**, 459–464 (2001).
- O'Hagan, R. C. et al. Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell* **2**, 149–155 (2002).
- Casanovas, O., Hager, J. H., Chun, M. G. & Hanahan, D. Incomplete inhibition of the Rb tumor suppressor pathway in the context of inactivated p53 is sufficient for pancreatic islet tumorigenesis. *Oncogene* **24**, 6597–6604 (2005).
- Bardeesy, N. et al. Both p16Ink4a and the p19Arf-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proc. Natl Acad. Sci. USA* **103**, 5947–5952 (2006).
- Pelham, R. J. et al. Identification of alterations in DNA copy number in host stromal cells during tumor progression. *Proc. Natl Acad. Sci. USA* **103**, 19848–19853 (2006).
- Kim, M. et al. Comparative oncogenomics identifies *NEDD9* as a melanoma metastasis gene. *Cell* **125**, 1269–1281 (2006).
- Zender, L. et al. Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* **125**, 1253–1267 (2006).
- O'Neil, J. et al. Activating *Notch1* mutations in mouse models of T-ALL. *Blood* **107**, 781–785 (2005).
- Sweet-Cordero, A. et al. Comparison of gene expression and DNA copy number changes in a murine model of lung cancer. *Genes Chromosom. Cancer* **45**, 338–348 (2006).
- Artandi, S. E. et al. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**, 641–645 (2000).
- Zhu, C. et al. Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell* **109**, 811–821 (2002).
- Lang, G. A. et al. Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell* **119**, 861–872 (2004).
- Olive, K. P. et al. Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119**, 847–860 (2004).
- Hingorani, S. R. et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
- Qi, L., Strong, M. A., Karim, B. O., Huso, D. L. & Greider, C. W. Telomere fusion to chromosome breaks reduces oncogenic translocations and tumour formation. *Nature Cell Biol.* **7**, 706–711 (2005).
- Qi, L. et al. Short telomeres and ataxia-telangiectasia mutated deficiency cooperatively increase telomere dysfunction and suppress tumorigenesis. *Cancer Res.* **63**, 8188–8196 (2003).
- Wong, K. K. et al. Telomere dysfunction and Atm deficiency compromises organ homeostasis and accelerates ageing. *Nature* **421**, 643–648 (2003).
- Shiloh, Y. & Kastan, M. B. ATM: genome stability, neuronal development, and cancer cross paths. *Adv. Cancer Res.* **83**, 209–254 (2001).
- Liyanage, M. et al. Abnormal rearrangement within the α/δ T-cell receptor locus in lymphomas from *Atm*-deficient mice. *Blood* **96**, 1940–1946 (2000).
- Rudolph, K. L., Millard, M., Bosenberg, M. W. & DePinho, R. A. Telomere dysfunction and evolution of intestinal carcinoma in mice and humans. *Nature Genet.* **28**, 155–159 (2001).
- Windle, B., Draper, B. W., Yin, Y. X., O'Gorman, S. & Wahl, G. M. A central role for chromosome breakage in gene amplification, deletion formation, and amplicon integration. *Genes Dev.* **5**, 160–174 (1991).
- Radtke, F., Wilson, A., Mancini, S. J. & MacDonald, H. R. Notch regulation of lymphocyte development and function. *Nature Immunol.* **5**, 247–253 (2004).
- Ellisen, L. W. et al. *TAN-1*, the human homolog of the *Drosophila Notch* gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell* **66**, 649–661 (1991).
- Schroeter, E. H., Kisslinger, J. A. & Kopan, R. Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain. *Nature* **393**, 382–386 (1998).
- O'Hagan, R. C. et al. Array comparative genome hybridization for tumor classification and gene discovery in mouse models of malignant melanoma. *Cancer Res.* **63**, 5352–5356 (2003).
- Shigeno, K. et al. Disease-related potential of mutations in transcriptional cofactors CREB-binding protein and p300 in leukemias. *Cancer Lett.* **213**, 11–20 (2004).
- Winandy, S., Wu, P. & Georgopoulos, K. A dominant mutation in the *Ikaros* gene leads to rapid development of leukemia and lymphoma. *Cell* **83**, 289–299 (1995).
- Graux, C. et al. Fusion of *NUP214* to *ABL1* on amplified episomes in T-cell acute lymphoblastic leukemia. *Nature Genet.* **36**, 1084–1089 (2004).
- Orlicky, S., Tang, X., Willems, A., Tyers, M. & Sicheri, F. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* **112**, 243–256 (2003).
- Mao, J. H. et al. *Fbxw7/Cdc4* is a p53-dependent, haploinsufficient tumour suppressor gene. *Nature* **432**, 775–779 (2004).
- Minella, A. C. & Clurman, B. E. Mechanisms of tumor suppression by the SCFFbw7. *Cell Cycle* **4**, 1356–1359 (2005).
- Sansal, I. & Sellers, W. R. The biology and clinical relevance of the *PTEN* tumor suppressor pathway. *J. Clin. Oncol.* **22**, 2954–2963 (2004).
- Suzuki, A. et al. High cancer susceptibility and embryonic lethality associated with mutation of the *PTEN* tumor suppressor gene in mice. *Curr. Biol.* **8**, 1169–1178 (1998).
- Yang, L. et al. Akt/protein kinase B signaling inhibitor-2, a selective small molecule inhibitor of Akt signaling with antitumor activity in cancer cells overexpressing Akt. *Cancer Res.* **64**, 4394–4399 (2004).
- Carrasco, D. R. et al. High-resolution genomic profiles define distinct clinicopathogenetic subgroups of multiple myeloma patients. *Cancer Cell* **9**, 313–325 (2006).
- Tonon, G. et al. High-resolution genomic profiles of human lung cancer. *Proc. Natl Acad. Sci. USA* **102**, 9625–9630 (2005).
- Futreal, P. A. et al. A census of human cancer genes. *Nature Rev. Cancer* **4**, 177–183 (2004).
- Rajagopalan, H. et al. Inactivation of *hCDC4* can cause chromosomal instability. *Nature* **428**, 77–81 (2004).
- Kemp, Z. et al. *CDC4* mutations occur in a subset of colorectal cancers but are not predicted to cause loss of function and are not associated with chromosomal instability. *Cancer Res.* **65**, 11361–11366 (2005).
- Kwak, E. L. et al. Infrequent mutations of Archipelago (*hAGO*, *hCDC4*, *Fbw7*) in primary ovarian cancer. *Gynecol. Oncol.* **98**, 124–128 (2005).
- Sharma, V. M., Draheim, K. M. & Kelliher, M. A. The Notch1/c-Myc Pathway in T cell leukemia. *Cell cycle* **6**, 327–330 (2007).
- Chin, L. et al. p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. *Cell* **97**, 527–538 (1999).
- Aguirre, A. J. et al. High-resolution characterization of the pancreatic adenocarcinoma genome. *Proc. Natl Acad. Sci. USA* **101**, 9067–9072 (2004).
- Mansour, M. R., Linch, D. C., Foroni, L., Goldstone, A. H. & Gale, R. E. High incidence of Notch-1 mutations in adult patients with T-cell acute lymphoblastic leukemia. *Leukemia* **20**, 537–539 (2006).
- Olszen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
- Davies, H. et al. Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res.* **65**, 7591–7595 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Zhang, A. Yu and K. Marmon for excellent mouse husbandry and care, and C. Greenman and E. Pleasance for helpful discussion on statistical analyses. R.S.M. was supported by the Damon Runyon Cancer Research Foundation. P.J.C. was supported by the Kay Kendall Leukaemia Fund, and B.C. is supported by a grant from GlaxoSmithKline. K.K.W. was supported by an NIH award. M.R.S. and P.A.F. are supported by the Wellcome Trust. L.C. and R.A.D. are supported by NIH grants, LeBow Fund to Cure Myeloma, the Chris Elliot Foundation, and the Center for Applied Cancer Science of the Belfer Institute for Innovative Cancer Science. R.A.D. is an Ellison Foundation for Medical Research Senior Scholar and an American Cancer Society Research Professor.

Author Contributions R.S.M., B.C., P.J.C. and B.F. performed the experiments and contributed equally as first authors. M.R.S., L.C., P.A.F. and R.A.D. supervised experiments and contributed equally as senior authors. R.S.M. and R.A.D. generated and characterized the instability mouse model. B.F. and L.C. conducted the oncogenomic analyses. B.C., P.J.C., M.R.S. and P.A.F. provided the re-sequencing analyses. A.P., J.O., A.G., E.I., I.P., E.L., V.M., S.J., K.M., S.Z., S.E., C.S., G.H., C.B., E.S.M., R.W., O.K., C.N., M.M. and V.D. performed experiments. A.G., L.F., A.K.F., A.H.G., J.M.R. and A.T.L. contributed patient samples and clinical data. K.K.W., J.A. and A.T.L. coordinated experiments. Y.A.W. contributed to the writing of the manuscript.

Author Information All microarray data are available at the Gene Expression Array Omnibus website (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE7615. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.A.D. (ron_depinho@dfci.harvard.edu).

METHODS

Mice. *Terc Atm* and *Terc Trp53* deficient mice have been described previously^{21,46}. Mice were interbred and maintained in pathogen-free conditions at Taconic Farms and Dana-Farber Cancer Institute, monitored for signs of ill-health every other day, and euthanized and necropsied when moribund. Mice found dead were necropsied specifically for signs of lymphoma. All manipulations were performed with IACUC approval.

Tumour characterization and sample preparation. Tumours harvested from TKO mice were partitioned for DNA, RNA and protein extraction, for histology (Brigham and Women's Hospital), and for *in vitro* culture. Tumour cells were cultured in RPMI with 50 µM β-mercaptoethanol, 10% CosmicCalf serum (HyClone), 0.5 ng ml⁻¹ IL-2 (Peprotech), and 4 ng ml⁻¹ IL-7 (Peprotech). For FACS analysis, cells were immunostained with CD4, CD8, CD3 and B220/CD45R antibodies (eBioscience) and scanned on a FACScalibur (BD Biosciences). For SKY analyses, metaphases were obtained from colcemid-treated cells incubated in 105 mM KCl hypotonic buffer for 15 min before fixation in 3:1 methanol-acetic acid. Spectral karyotyping was done using the SkyPaint Kit and SkyView analytical software (Applied Spectral Imaging) according to manufacturer's protocols. Chromosome aberrations were defined in accordance with the Committee on Standard Genetic Nomenclature for Mice.

DNA was prepared with the PureGene kit (Gentra Systems). Real-time PCR was performed with a Quantitect SYBR green kit (Qiagen) using 2 ng DNA from each tumour run in triplicate and the primers listed in Supplementary Table 7, on Applied Biosystems or Stratagene MX3000 real-time thermocyclers; copy number average of two runs was calculated using the standard curve method. RNA was extracted with Trizol (Invitrogen) according to the manufacturer's instructions, then digested with RQ1 DNase (Promega) and purified through RNA purification columns (Gentra). Reverse transcription was performed with oligo dT primers with the Oligoscript kit (Qiagen). Protein lysates were prepared by dis-aggregation in lysis buffer (Cell Signaling Technology) followed by bath sonication for 30 s. Clarified lysates were quantified with BioRad protein assay according to manufacturer's instructions, separated on 4–12% NuPage gels (Invitrogen), and transferred to PVDF membranes (Immobilon). Immunoblotting was performed with antibodies against PTEN, AKT, phospho-AKT, NOTCH1, activated NOTCH1 Val1744 (all from Cell Signaling Technology), and tubulin (Sigma) and developed with HRP-labelled secondary antibodies (Pierce) and enhanced chemiluminescent substrate.

Human T-ALL samples and cell lines. T-ALL cell lines were subjected to both array-CGH and/or re-sequencing, as indicated (Supplementary Table 2a).

Two cohorts of clinical human T-ALL samples were used. A cohort of 8 clinical samples (Supplementary Table 2b), subjected to array-CGH, was obtained with informed consent and IRB approval at time of diagnosis from paediatric patients with T-ALL treated on the Dana-Farber Cancer Institute study 00-001. An independent series of 38 clinical specimens (Supplementary Table 2c), used for re-sequencing, was collected at presentation from 8 children and adolescents diagnosed at the Royal Free Hospital, London, and 30 adult patients enrolled in the MRC UKALL-XII trial. Informed consent was obtained from the patients (if over 18 yr of age) or their guardians (if under 18 yr) with Ethics Committee approval.

Array-CGH profiling and analyses. Array-CGH: genomic DNA processing, labelling and hybridization to Agilent CGH 60-mer oligo arrays were performed as per the manufacturer's protocol (<http://www.home.agilent.com/agilent/home.jspx>). Murine tumours were profiled against matched normal DNA or, when not available, pooled DNA of matched strain background. Labelled DNAs were hybridized onto 44K or 244K microarrays for mouse, and 22K, 44K, or 244K microarrays for human (detailed feature information for each at www.agilent.com). Profiles generated on 244K arrays were extracted for the same 42K probes on the 44K microarrays to allow combination of profiles generated on the different platforms. Fluorescence ratios of scanned images were normalized and averaged from two pairs (dye swap), and copy number profile was generated by Circular Binary Segmentation, which determines significance of change points in raw data through permutation⁴⁹.

Definition of MCRs: minimal common region (MCR) definition was described previously^{39,47}. Briefly, a 'segmented' data set was determined from uniform copy number segment boundaries, then replacing raw log₂ ratio for each probe by the mean log₂ ratio of the segment containing the probe. Thresholds for copy number alterations were set at log₂ = ±0.4 and ±0.6 for 22K and 44K arrays, respectively, with higher thresholds for 44K profiles adjusting for differences in signal:noise detection. MCRs required at least one sample to show an extreme CNA event, (log₂ ratio ≥ ±0.60 and ±0.75 for 22K and 44K profiles, respectively, and the width less than 10 Mb).

Homologue mapping: genes within mouse TKO MCRs were used to obtain their human homologue on the basis of the NCBI homologene website (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene>). For cancer gene

mapping, mouse homologues were obtained on the basis of the Sanger Cancer Gene Census⁴¹ (<http://www.sanger.ac.uk/genetics/CGP/Census>).

Permutation testing: permutation testing was conducted for statistical significance of MCR overlap by randomly generating a simulated mouse genome containing the same number and sizes of MCRs in the corresponding chromosomes as the actual TKO genome; a similar set was created for each human cancer data set. The number of overlapping MCRs between mouse and each human genome was calculated and stored, and repeated 10,000 times independently for amplifications and deletions. The *P* value for significance of overlap for each was calculated by dividing the frequency of randomly achieved overlap by 10,000.

Mutation screening. Mutational status was established by denaturing high-performance liquid chromatography⁴⁸ and by bidirectional sequencing. Genomic DNA was extracted using the Qiagen genomic purification kit. PCR amplification and direct sequencing were done as previously described⁵⁰. Sequence traces were analysed using manual and software-based analyses. All variants were confirmed by bidirectional sequencing of a second independently amplified PCR product.

Expression profiling. Biotinylated target complementary RNA was generated from total sample RNA and hybridized to 44K Mouse Development Oligo Microarrays (Agilent) against normal control murine thymus RNA according to the manufacturer's protocols. Expression values for each gene were mapped to genomic positions on the basis of National Center for Biotechnology Information Build 34 of the mouse genome.

Drug sensitivity. Cells (20,000) were plated in triplicate in 96-well format and incubated in media with varying doses of triciribine (BioMol) or vehicle (DMSO; Sigma) for 2 days, after which growth was quantified using the AqueousOne Cell Titer System (Promega). Relative cell growth was plotted versus equivalent amount of DMSO alone as control, and 3–5 replicates were performed for each.

ARTICLES

Gene-specific control of inflammation by TLR-induced chromatin modifications

Simmie L. Foster^{1*}, Diana C. Hargreaves^{1*} & Ruslan Medzhitov¹

Toll-like receptors (TLRs) induce a multi-component inflammatory response that must be tightly regulated to avoid tissue damage. Most known regulatory mechanisms target TLR signalling pathways and thus broadly inhibit multiple aspects of the inflammatory response. Given the functional diversity of TLR-induced genes, we proposed that additional, gene-specific regulatory mechanisms exist to allow individual aspects of the TLR-induced response to be differentially regulated. Using an *in vitro* system of lipopolysaccharide tolerance in murine macrophages, we show that TLR-induced genes fall into two categories on the basis of their functions and regulatory requirements. We demonstrate that representatives from the two classes acquire distinct patterns of TLR-induced chromatin modifications. These gene-specific chromatin modifications are associated with transient silencing of one class of genes, which includes pro-inflammatory mediators, and priming of the second class, which includes antimicrobial effectors. These findings illustrate an adaptive response in macrophages and reveal component-specific regulation of inflammation.

Inflammation is a complex response to infection and tissue injury¹. Toll-like receptors (TLRs) have a critical role in the inflammatory response to infection. Thus, bacterial lipopolysaccharide (LPS) signals through TLR4 and is one of the most potent inducers of inflammation². Because the inflammatory response causes marked changes in tissue physiology, dysregulated inflammation can lead to a variety of pathological conditions, including septic shock, autoimmunity, atherosclerosis and metabolic syndrome^{3,4}. Accordingly, the inflammatory response must be tightly regulated and indeed, multiple regulatory mechanisms control the extent and duration of TLR-induced inflammation. These include the inhibition of TLR signalling by inducible negative regulators, production of anti-inflammatory cytokines and alterations of the TLR signalling complex⁵. Collectively, these mechanisms contribute to the phenomenon of 'LPS tolerance': the transient unresponsiveness of cells or organisms to repeated or prolonged stimulation with LPS^{6–9}.

LPS tolerance has traditionally been viewed as a hyporesponsive state of macrophages resulting from receptor desensitization^{10–15}. However, TLRs induce expression of hundreds of genes with different functions¹⁶ and therefore different regulatory requirements. Thus, it is unlikely that all TLR-induced genes are controlled solely at the signalling level, as this would not discriminate between gene subsets with distinct functions. For example, not all TLR-induced genes have the potential to cause tissue damage. These include genes encoding antimicrobial effectors that are essential for the early host defence from infection. Even transient disruption of TLR-induced expression of these genes would leave the host immunocompromised.

We reasoned, therefore, that TLR-induced genes with different biological functions should have distinct requirements for regulation. Specifically, genes encoding pro-inflammatory mediators should be transiently inactivated in tolerant macrophages to limit tissue damage. On the other hand, genes encoding antimicrobial effectors and other proteins that do not negatively affect tissue physiology should remain inducible even after repeated stimulation of TLRs to provide continuous protection from infection. Here we demonstrate that this is indeed the case; we have found that TLR4-induced genes

fall into two categories on the basis of their functions and regulatory requirements. Because genes from both classes are induced by the same receptor, their expression is regulated by gene-specific rather than signal-specific mechanisms. We further show that there is an adaptive element to the innate immune response, and that this adaptation is based on epigenetic mechanisms.

Identification of two classes of TLR4-induced genes

We first tested the hypothesis that induction of LPS tolerance would selectively inhibit expression of pro-inflammatory genes, whereas genes encoding antimicrobial effectors would remain inducible. Consistent with previous studies, induction of the pro-inflammatory cytokine interleukin-6 (IL-6) was abolished after an initial LPS stimulation, and this state of LPS tolerance persisted for 24–48 h (Fig. 1a, b)⁸. However, an antimicrobial gene, *Cnlp* (cathelicidin-related antimicrobial peptide; also called *Camp*), remained inducible in tolerant macrophages under the same conditions (Fig. 1c). We next performed a microarray analysis comparing unstimulated macrophages, naive macrophages stimulated with LPS, or tolerant macrophages re-stimulated with LPS (Supplementary Table 1). As expected, several hundred genes induced during the first stimulation were either not re-induced or induced to a much lesser degree by a second stimulation at the 24-h time point, when macrophages exhibit maximal tolerance (Fig. 1d). Notably, a second group, also comprising several hundred genes induced during the first LPS stimulation, was induced at equal or greater levels after the stimulation of tolerant macrophages (Fig. 1d). We have categorized these genes into two classes: class 'tolerizable' (genes not inducible in tolerant macrophages, class T) and class 'non-tolerizable' (genes inducible in tolerant macrophages, class NT). We verified these distinct expression patterns by analysing at least ten representative genes with known function from each category by quantitative polymerase chain reaction (qPCR) (Fig. 1e). Although both T and NT classes contained genes from multiple functional categories (Supplementary Fig. 1 and Supplementary Table 2), the biological significance of differential regulation is most obvious for pro-inflammatory (class

¹Howard Hughes Medical Institute and Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut 06405, USA.

*These authors contributed equally to this work.

T) and antimicrobial (class NT) genes. Therefore, for further analyses we chose several genes that belonged to these functional groups and other select genes that displayed robust regulation.

The glucocorticoid receptor (GR) is known to negatively regulate pro-inflammatory genes^{17,18}. We found that the GR agonist dexamethasone selectively inhibited the induction of several class T genes, but did not inhibit, and in some cases enhanced, the induction of several class NT genes (Fig. 1f and Supplementary Fig. 2a). This differential sensitivity to GR further supports the biological significance of distinction between the two classes of LPS-induced genes. One mechanism by which GR exerts gene-specific effects is the differential use of IRF3 as a co-activator for NF-κB-dependent transcription¹⁷. We found, however, that most genes in both classes were dependent on IRF3 to varying degrees (data not shown).

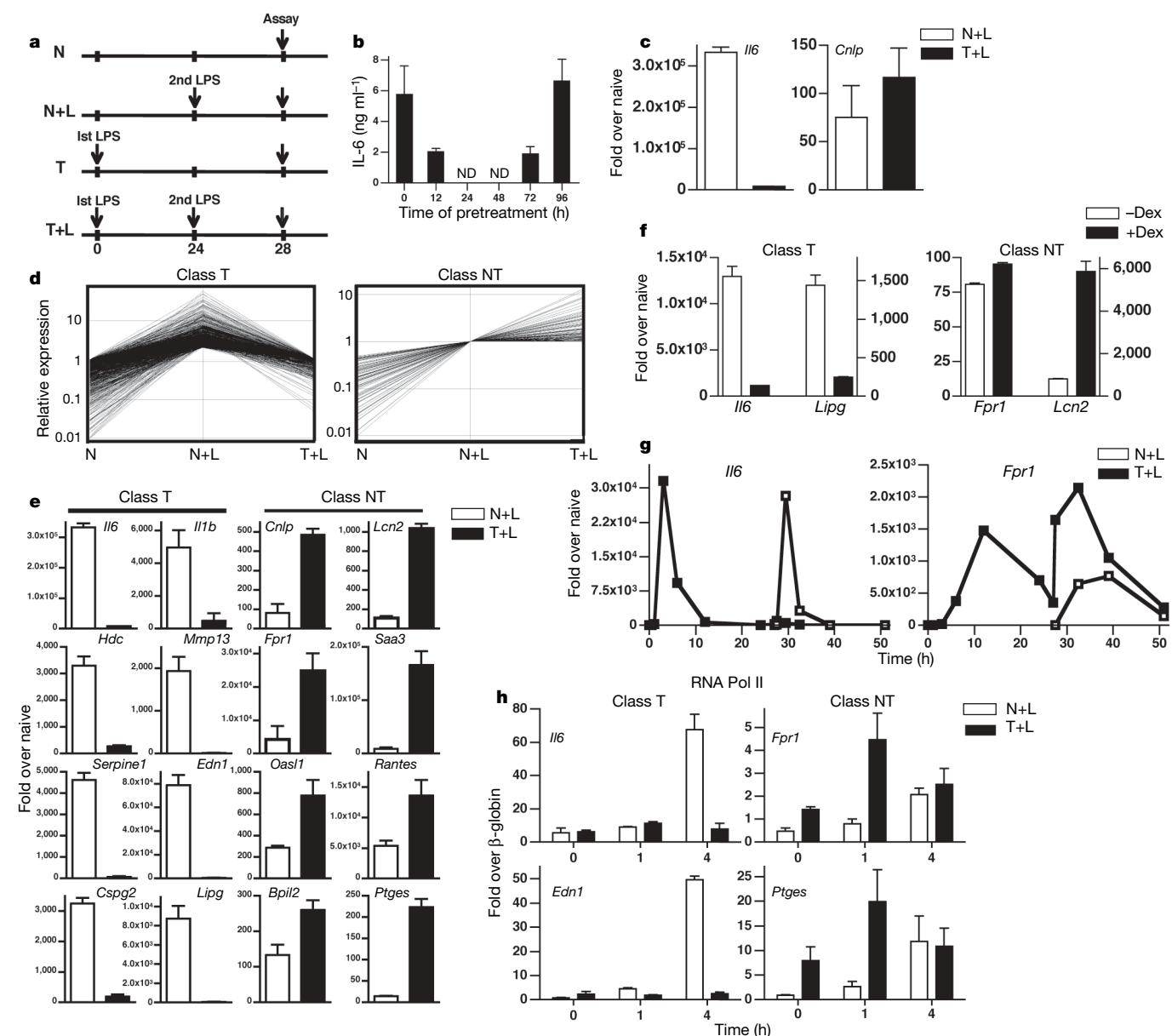


Figure 1 | Identification of class T and class NT genes. **a**, BMMFs were left untreated (naive, N) or stimulated with 100 ng ml^{-1} LPS for 24 h (tolerant, T), washed with PBS and given media (N, T) or 10 ng ml^{-1} LPS (N+L, T+L). Cells were assayed at time points post-stimulation. **b**, BMMFs were pretreated with 100 ng ml^{-1} LPS for the indicated times and restimulated with 10 ng ml^{-1} LPS for 24 h. Supernatants were analysed for IL-6 by ELISA. **c–e**, BMMFs were stimulated as described in **a**. RNA was harvested after 4 h and analysed by Affymetrix genechip (**d**) or by RT-qPCR (**c, e**). **f**, BMMFs

were next asked whether the two classes of genes were regulated at the level of transcription. First, we analysed expression kinetics of several class T and class NT genes. Although *Il6* (class T) was not re-inducible in tolerant macrophages, *Fpr1* (formyl peptide receptor 1; class NT) was induced to a greater extent and with faster kinetics in tolerant cells than in naive cells (Fig. 1g). Several other NT genes also appeared to be ‘primed’ by the first stimulus (Supplementary Fig. 2b). We measured the stability of messenger RNAs of several T and NT genes and found that several mRNAs from both classes had an overlapping range of stabilities that did not correlate with their expression in tolerant macrophages (Supplementary Fig. 3a). Finally, we performed a chromatin immunoprecipitation (ChIP) assay for RNA polymerase II (Pol II). Although Pol II was inducibly recruited to both classes of promoters in naive macrophages, in

were stimulated with 100 ng ml^{-1} LPS +/− Dex for 6 h and analysed by RT-qPCR. **g**, BMMFs were stimulated with (black squares) or without (white squares) 100 ng ml^{-1} LPS, washed at 24 h and restimulated with 10 ng ml^{-1} LPS. RNA was analysed by RT-qPCR at indicated times. **h**, BMMFs were stimulated as described in **a** and analysed by ChIP (RNA Pol II). **b, c, e–h**, Data are representative of 2 or more independent experiments. Data show mean ± s.e.m. from triplicate values. **d**, Data represents the average expression of two independent experiments.

tolerant macrophages Pol II was only recruited to the class NT promoters and this occurred with faster kinetics than in naive macrophages (Fig. 1h). These results demonstrate that class NT genes, unlike class T genes, are indeed transcriptionally inducible in tolerant macrophages.

We next asked whether the two classes of genes are regulated by the same signals. Consistent with previous reports^{11–15}, we found that the activation of NF-κB and mitogen-activated protein kinases (MAPK) was deficient in tolerant macrophages (Supplementary Fig. 4a). MAPK inhibitors blocked the induction of several class T and NT genes in naive macrophages (Supplementary Fig. 4b), as well as the induction of class NT genes in tolerant cells (Supplementary Fig. 4c), suggesting that the same signalling pathways contribute to class NT gene induction in naive and tolerant macrophages. Class T and NT genes were induced similarly by sub-optimal doses of LPS in naive macrophages (Supplementary Fig. 3b), suggesting that class NT genes are not intrinsically more sensitive to LPS. There was no correlation between interferon (IFN)-α/β receptor dependence and class T and NT gene expression, indicating that class NT genes are not regulated through IFN-α/β feedback (data not shown). Finally, class

NT genes were not induced by conditioned media from LPS-stimulated macrophages, suggesting that class NT gene induction is not the result of positive feedback of any other secreted factor (Supplementary Fig. 3c).

Collectively, these results indicate that even attenuated TLR4 signalling in tolerant macrophages is sufficient for the induction of class NT genes, but not class T genes.

Distinct histone modifications at T and NT promoters

These results suggested that the two classes of LPS-induced genes are differentially regulated by gene-specific characteristics. Gene-specific regulation occurs at the level of chromatin and includes nucleosome remodelling and covalent histone modifications^{19,20}. Histone acetylation is a positive mark associated with transcriptionally active chromatin, whereas deacetylated histones are found in closed, inactive chromatin^{21–23}. Although promoters of both classes were inducibly acetylated at histone H4 in naive macrophages, only histones at class NT promoters were re-acetylated after stimulation of tolerant macrophages (Fig. 2a). The kinetics of induction of histone acetylation at class NT promoters in tolerant macrophages mirrored gene

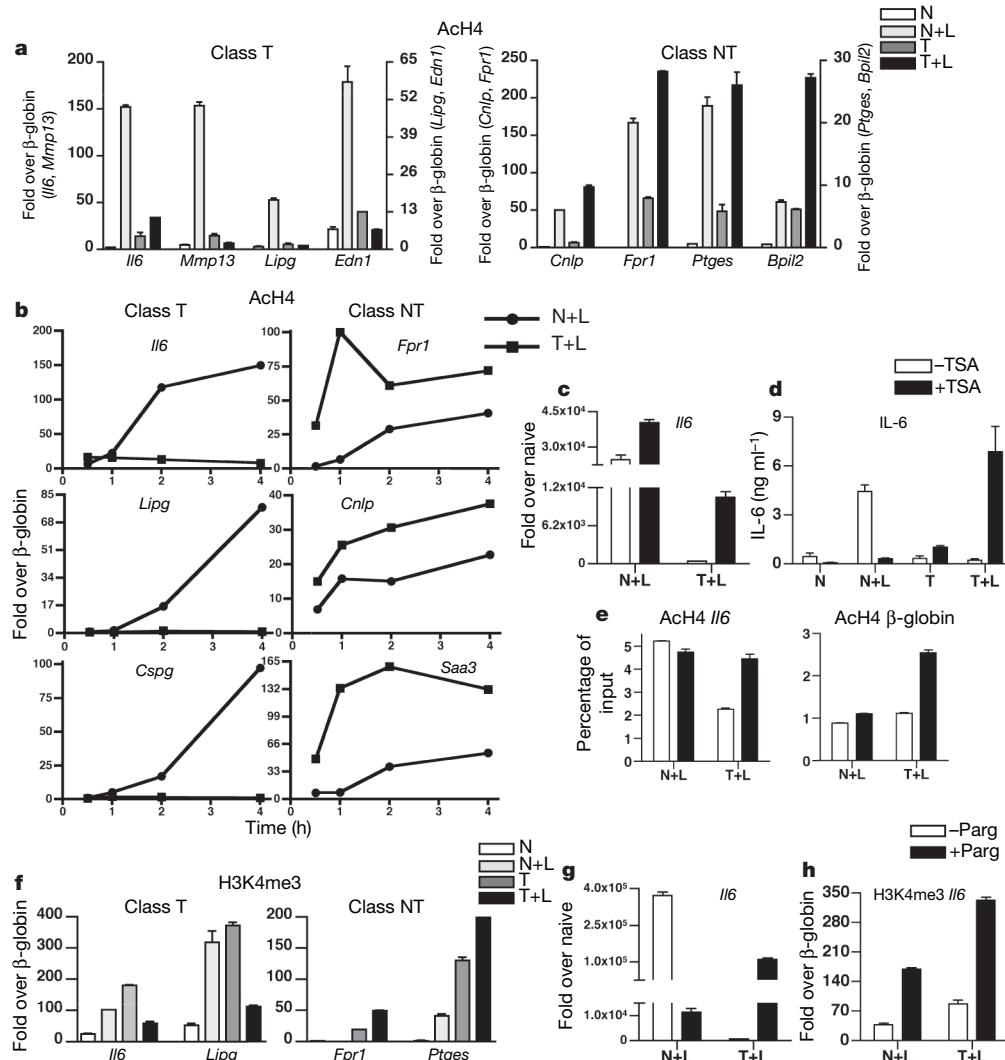


Figure 2 | Histone modifications are differentially regulated at class T and NT promoters. Naive and tolerant BMMφs (N, T) were stimulated with LPS for **a**, 3 h (N+L) or 1 h (T+L) or **b**, the indicated times and analysed by ChIP (AcH4). **c–e**, Naive BMMφs were stimulated with LPS (N+L, white bars) or LPS+TSA (N+L, black bars). Tolerant BMMφs (T+L) were prepared as above (white bars) or with TSA and stimulated with LPS+TSA (black bars). RT-qPCR (**c**), ELISA (**d**) and ChIP (AcH4) (**e**) were performed at 4 h (**c, e**) or

24 h (**d, f**). **f**, Naive and tolerant BMMφs (N, T) were stimulated with LPS for 3 h (N+L, T+L) and analysed by ChIP (H3K4me3). **g, h**, Naive and tolerant BMMφs were prepared with pargyline as in **c** and **e**. RT-qPCR (**g**) and ChIP (H3K4me3) (**h**) were performed at 4 h. **a–h**, Data are representative of 3 or more independent experiments; shown are mean ± s.e.m. from triplicate values.

expression: acetylation persisted at a higher level relative to unstimulated cells, and was re-induced by stimulation of tolerant cells, in some cases with faster kinetics and to a greater extent (Fig. 2b and Supplementary Fig. 2c). Thus, class T and class NT genes exhibit distinct patterns of inducible histone acetylation that correspond to their transcriptional activity.

We next stimulated macrophages in the presence of trichostatin A (TSA), a histone deacetylase inhibitor, and measured the induction of *Il6*. As others have shown previously, TSA inhibited *Il6* expression in naive macrophages stimulated with LPS, presumably by affecting the acetylation of transcription factors^{24,25}. However, inhibition of histone deacetylases during the first LPS stimulation of naive macrophages reversed silencing of *Il6* and several other class T genes in tolerant macrophages (Fig. 2c, d and data not shown). TSA treatment did not affect NF-κB or MAPK activation (Supplementary Fig. 5a), but did result in an increased level of histone acetylation at the *Il6* promoter in tolerant macrophages (Fig. 2e). Thus, selective histone deacetylation contributes to the silencing of class T genes in tolerant macrophages.

Trimethylation of histone H3 at lysine 4 (H3K4me3) also marks transcriptionally active genes^{26–29}. We found that H3K4 trimethylation was induced in naive macrophages at both classes of promoters (Fig. 2f). Interestingly, following LPS stimulation of tolerant macrophages, this modification was rapidly and selectively lost at class T promoters, but was maintained at class NT promoters (see Fig. 2f, and below). Treatment of macrophages with pargyline, an inhibitor of H3K4 demethylase LSD1 (ref. 30), prevented *Il6* silencing in tolerant macrophages (Fig. 2g), and maintained H3K4me3 levels at the *Il6* promoter (Fig. 2h). Pargyline treatment did not affect NF-κB and MAPK activation by LPS (Supplementary Fig. 5b). Thus two types of positive histone modifications, H4 acetylation and H3K4 trimethylation, are selectively lost at the class T promoters; inhibiting their loss prevents silencing of the class T genes in tolerant macrophages.

Nucleosome remodelling at T and NT promoters

We next examined the recruitment of two ATP-dependent chromatin remodelling complexes, Brg1 and Mi-2β, to class T and class NT promoters in naive and tolerant macrophages²⁰. Following LPS stimulation, Brg1 was recruited to both classes of promoters in naive macrophages, but was only recruited to class NT promoters in tolerant macrophages (Fig. 3a). Similar to the recruitment of Pol II, Brg1 was recruited with faster kinetics to class NT promoters in tolerant macrophages (Fig. 3a, *Fpr1*). As shown previously, Mi-2β was recruited simultaneously with Brg1 in all cases (data not shown)³¹.

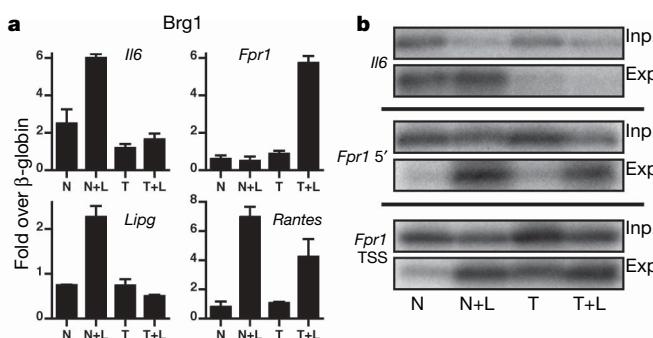


Figure 3 | Chromatin remodelling is differentially regulated at class T and NT promoters. **a**, Naive and tolerant BMMφs were left untreated (N, T) or were stimulated with LPS (N+L, T+L) for 3 h. Cells were fixed with DMA-formaldehyde and analysed by ChIP (Brg1). **b**, BMMφs were stimulated as in **a** for 3 h (*Il6* (N+L, T+L), *Fpr1* (T+L)) or 8 h (*Fpr1* (N+L)) and analysed by REA/LM-PCR. Input (Inp.) and experimental (Exp.) amplification products are shown. **a, b**, Data are representative of 3 or more independent experiments; shown are mean ± s.e.m. from triplicate values.

Consistent with the pattern of Brg1 recruitment, *Il6* and *Fpr1* promoters became nuclelease accessible in naive macrophages stimulated with LPS. However, following stimulation of tolerant macrophages, *Il6* was inaccessible, whereas *Fpr1* was highly accessible at even earlier time points than in naive macrophages (Fig. 3b, *Fpr1* TSS). These data demonstrate that in tolerant macrophages chromatin remodelling is inhibited at class T promoters, whereas class NT promoters are both stably and inducibly remodelled.

TLR4-induced transcripts regulate T and NT genes

We next investigated whether gene products induced by LPS in naive macrophages are required for the silencing of class T genes and the priming of class NT genes in tolerant macrophages. The transcription elongation inhibitor DRB was used to block LPS-induced transcription in naive macrophages. As expected, DRB inhibited the transcription of several class T genes in LPS-stimulated naive macrophages (Fig. 4a, b). Interestingly, DRB treatment of naive macrophages completely prevented silencing of *Il6* and several other class T genes in tolerant macrophages (Fig. 4b, c). Accordingly, LPS stimulation led to the recruitment of Pol II, NF-κB and Brg1 to the *Il6* promoter in DRB pre-treated, but not in untreated, tolerant macrophages, and this recruitment occurred with faster kinetics than in naive macrophages (Fig. 4d). Similarly, we found increased H4 acetylation and sustained levels of H3K4me3 at the *Il6* and *Lipg* promoters in DRB-treated macrophages (Fig. 4e). Thus, DRB treatment of naive macrophages results in a reversal of tolerance in LPS-pretreated macrophages, as determined by the induction of positive histone modifications (H4 acetylation and H3K4 trimethylation), the recruitment of transcription factors and nucleosome remodelling complexes, and finally, the recruitment of the Pol II complex and transcriptional induction of class T genes.

Because negative transcriptional regulators induced by the first LPS stimulation seem to be responsible for silencing class T genes, we wondered if positive regulators similarly induced are responsible for priming class NT genes in tolerant macrophages. Therefore, we examined the effect of DRB on induction of class NT genes. Similarly to class T genes, class NT genes were inhibited by DRB treatment of naive macrophages stimulated with LPS (Fig. 4f). However, in contrast to the DRB-mediated reversal of silencing observed at class T genes, DRB pre-treatment prevented priming of class NT genes in tolerant macrophages. Both the magnitude of their transcriptional induction and the delayed kinetics of H3K4 trimethylation were similar to that seen in LPS-stimulated naive macrophages (Fig. 4f, g).

DRB treatment did not affect NF-κB and MAPK signalling in naive and tolerant macrophages or the induction of negative regulators of TLR signalling, such as Irak-M (Supplementary Fig. 5c, d). These controls confirmed that the effect of DRB was gene-specific, rather than signal-specific. In addition, these experiments demonstrate that reduced signalling in tolerant macrophages is sufficient for class T induction in the absence of negative regulators, illustrating that permissive chromatin structure, not signal strength or specificity, is the determining factor for differential gene induction.

Altered transcriptional requirements of NT genes

As discussed above, class NT genes are induced with enhanced kinetics and magnitude in tolerant macrophages, indicating that their transcriptional requirements change as a result of the first exposure to LPS (Fig. 1g). One characteristic of inducible genes that correlates with the kinetics of induction is the requirement for protein synthesis³¹. Primary response genes are not dependent on protein synthesis and are induced within 2 h of stimulation. Secondary response genes are induced with later kinetics and their transcription depends on protein synthesis. Both class T and class NT genes contained

primary and secondary genes. However, the change in kinetics of induction of class NT genes in naive and tolerant macrophages raised the possibility that those class NT genes that are secondary in naive macrophages may be converted into primary genes in tolerant macrophages. We therefore chose several secondary class NT genes and measured their induction in naive and tolerant macrophages in the presence of cycloheximide (CHX), a protein synthesis inhibitor. We found that although these genes were inhibited by CHX in naive macrophages, CHX treatment had no effect on the transcription of class NT genes in tolerant macrophages (Fig. 5a). Furthermore, Pol II recruitment to promoters of secondary class NT genes was CHX sensitive in naive macrophages, but CHX insensitive in tolerant macrophages (Fig. 5b). Similarly, the accessibility of the *Fpr1* promoter (class NT) was dependent on new protein synthesis in naive, but not tolerant, macrophages (Fig. 5c). These results indicate that secondary class NT genes are indeed converted into primary genes as a result of macrophage 'priming' by the first stimulation with LPS.

Discussion

TLR activation induces expression of hundreds of genes in macrophages. These fall into multiple functional categories, including inflammatory cytokines, chemokines, antimicrobial proteins and

peptides, tissue-repair and coagulation factors, and metabolic regulators. We reasoned that different components of the TLR-induced response should have different regulatory requirements that reflect their functions. We show here that TLR4-induced genes fall into two classes. On repeated exposure to LPS, one class of genes (class T, including inflammatory cytokines) is transiently silenced to prevent pathology associated with excessive inflammation. The second class of genes (class NT) includes antimicrobial effectors, which remain inducible to protect the host from infection.

Although our initial hypothesis was based on the functional distinction between pro-inflammatory and antimicrobial genes, a more comprehensive gene expression analysis revealed that other functional groups of LPS-induced genes belong to these two categories (Supplementary Tables 1, 2; Supplementary Fig. 1). Indeed, several groups have reported examples of anti-inflammatory gene induction in tolerant macrophages^{32–36}. In all cases, the classification of genes as class T or NT probably depends on whether persistent inducibility of a gene would be deleterious or advantageous, and thus reflects a general principle linking gene regulation with the function of the encoded products. Because both classes of genes are induced by the same receptor, their differential regulation occurs by gene-specific rather than signal-specific mechanisms, through chromatin modifications at the

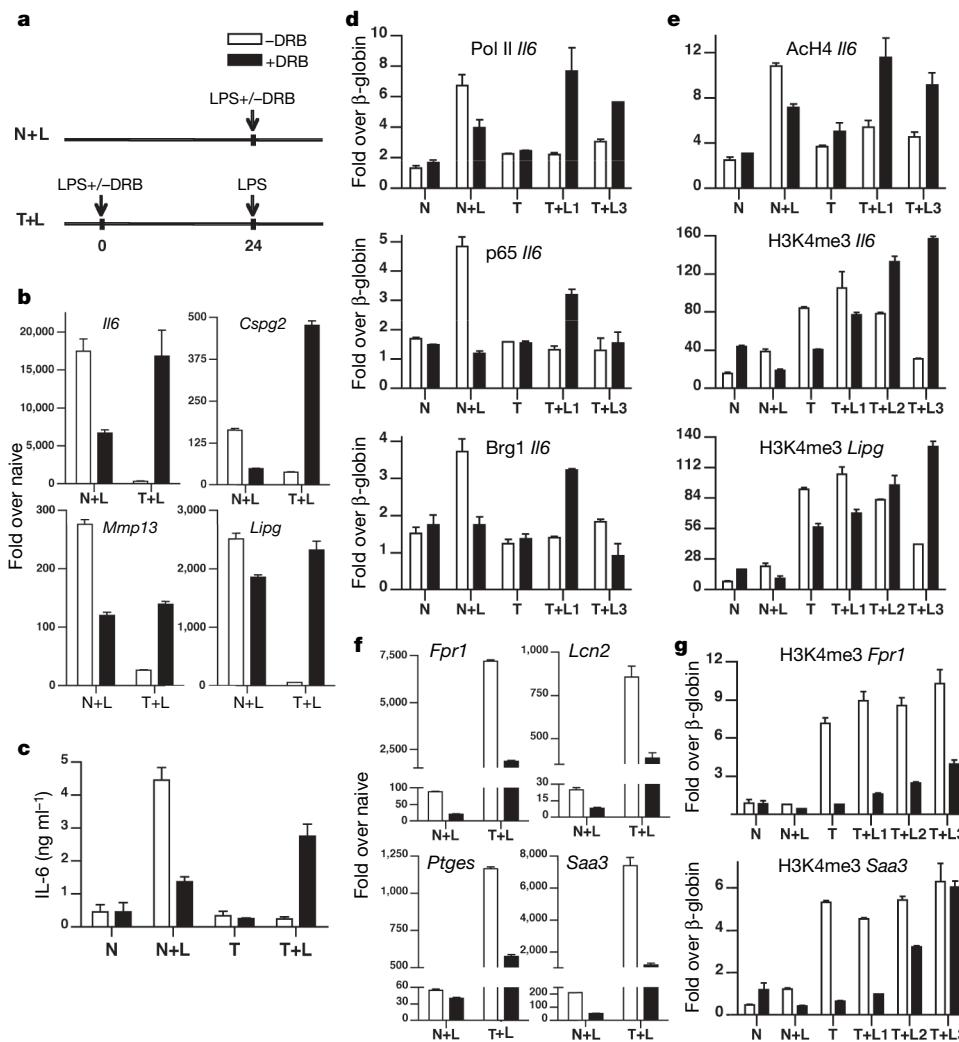


Figure 4 | Transcription of new genes contributes to the tolerant

signature. **a**, Naive BMMφs were left untreated (N), or were stimulated after 24 h with LPS (N+L, white bars) or LPS+DRB (N+L, black bars). Tolerant BMMφs were prepared with or without DRB, washed with PBS, then left untreated (T) or stimulated with LPS (T+L). In all panels, white and black bars indicate stimulation in the absence or presence of DRB, respectively.

b, f, g, RT-qPCR was performed 4 h after stimulation. **c**, Supernatants were analysed by ELISA for IL-6 24 h after stimulation. **d**, ChIP analysis was performed at 1 h (T+L1) or 3 h (N+L, T+L3) for RNA Pol II, p65 and Brd1. **e, g**, ChIP was performed at 1 h (T+L1), 2 h (T+L2) or 3 h (N+L, T+L3) for AcH4 or H3K4me3. **b–g**, Data are representative of 3 or more independent experiments; shown are mean \pm s.e.m. from triplicate values.

level of individual promoters. We speculate that similar mechanisms operate in other multi-component transcriptional response programmes.

Several mechanisms regulate the TLR-induced inflammatory response and these collectively contribute to LPS tolerance *in vitro* and *in vivo*. Our results demonstrate that selective and transient silencing of inflammatory genes at the level of chromatin plays a critical role in LPS tolerance.

We show that gene products (presumably transcriptional regulators) induced by the first LPS stimulation of naive macrophages are required for the silencing of class T genes and the priming (enhanced inducibility) of class NT genes in tolerant macrophages (Fig. 6). In contrast to the long-term epigenetic memory that propagates inheritance of gene expression through cell divisions, this is an example of transient gene silencing in terminally differentiated macrophages. Stable gene silencing is associated with histone H3 methylation at H3K9 and H3K27 (ref. 37). We did not detect these modifications at class T genes in tolerant macrophages (data not shown), suggesting that this type of transient gene silencing occurs by a different mechanism.

We found that following the initial exposure to LPS, class NT genes become modified so that their induction by a second LPS stimulation occurs with faster kinetics and an increased magnitude. This transcriptional memory may constitute an adaptive component of the innate immune response; as many NT genes encode antimicrobial effectors, this enhanced response would increase the efficiency of innate host defence.

A recent study found that primary and secondary genes differ in their requirement for inducible nucleosome remodelling mediated by Brg1 (ref. 31). Primary response gene products recruit Brg1 to the promoters of secondary response genes, which explains why secondary response genes require protein synthesis³¹. We found that the transcriptional requirements for class NT genes change during LPS stimulation, such that secondary NT genes are converted into primary genes in tolerant macrophages. This conversion may reflect either inducible recruitment of pre-made primary gene products to the promoters, or persistent changes in chromatin structure initiated

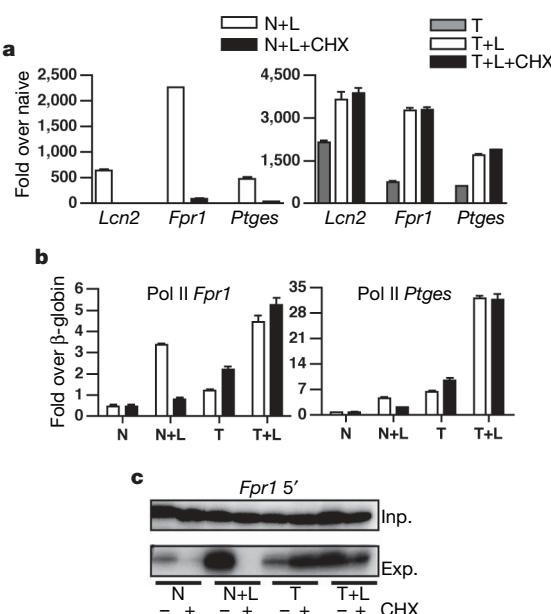


Figure 5 | Class NT genes have different transcriptional requirements in naive and tolerant macrophages. **a**, Naive BMMΦs were stimulated with LPS (white bars) or LPS+CHX (black bars) for 6 h. Tolerant BMMΦs were left untreated (grey bars) or stimulated with LPS (white bars) or LPS+CHX (black bars) for 2 h. RT-qPCR was performed. **b, c**, BMMΦs were stimulated as in **a**, but cells were analysed by ChIP (RNA Pol II) (**b**) or REA/LM-PCR (**c**) after 3 h. **a–c**, Data are representative of 2 or more independent experiments; **a, b**, shown are mean \pm s.e.m. from triplicate values.

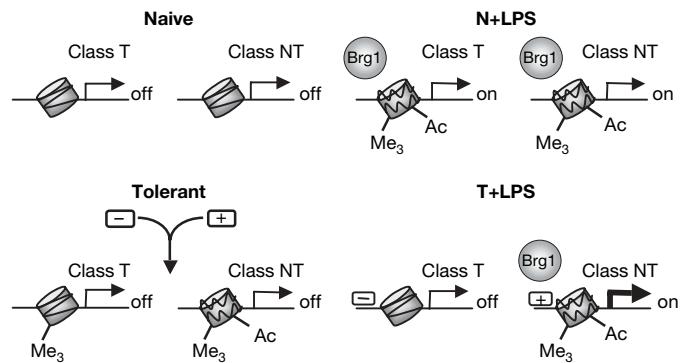


Figure 6 | Model for gene-specific regulation of class T and NT genes. Following LPS stimulation of naive macrophages, class T and NT promoters exhibit transcription factor recruitment, increased histone acetylation, H3K4 trimethylation, and chromatin remodelling. In tolerant macrophages, H3K4 trimethylation is high at both classes of promoters, and class NT promoters exhibit increased levels of histone acetylation and accessibility. Following stimulation with LPS, class T promoters remain deacetylated and inaccessible, whereas class NT promoters become even more acetylated and accessible, this time with faster kinetics. TLR4-induced negative and positive factors contribute to the silencing of class T promoters and the priming of class NT promoters in tolerant macrophages.

by primary gene products during the first stimulation. Regardless, this finding explains why the induction of class NT genes is qualitatively and quantitatively different in naive and tolerant macrophages.

Collectively, these results indicate that gene products induced by LPS in naive macrophages differentially modify chromatin at class T and class NT promoters to silence the former and to prime the latter for their differential regulation by a second LPS stimulation.

Safe manipulation of the innate immune response and inflammation has been problematic because most known therapeutic agents inhibit the induction of both antimicrobial effectors and inflammatory cytokines. Here we have shown that the two types of TLR-induced responses can be dissociated and are differentially regulated, suggesting the existence of novel targets for selective control of inflammatory and antimicrobial responses.

METHODS SUMMARY

Bone marrow macrophage (BMMΦ) cultures. Reagents are described in Methods. Bone marrow progenitors were harvested from mice and cultured for 7 days on Petri dishes in M-CSF supplemented RPMI-1640. Cells were lifted with cold PBS and replated on tissue-culture treated plates. On day 8, macrophages were left untreated (naive, N) or stimulated with 100 ng ml⁻¹ LPS for 24 h (tolerant, T), washed twice with warm PBS and given fresh media (N, T) or 10 ng ml⁻¹ LPS (N+L, T+L). Where indicated, BMMΦ were treated with dexamethasone (Dex; 1 μM), actinomycin D (ActD; 10 μg ml⁻¹), TSA (50 nM), DRB (15 μM), pargyline (3 μM), polymyxin B sulphate (PB; 50 μg ml⁻¹) and CHX (100 μg ml⁻¹). The p38/JNK inhibitor cocktail included PD98059 (45 μM), SB203580 (2.5 μM), SB202190 (0.5 μM), PD169136 (5 μM) and JNK inhibitor II (500 nM).

ELISA. Described in Methods.

Reverse transcription and qPCR. Total RNA was isolated, reverse transcribed, and analysed in triplicate by qPCR. Expression was normalized to hypoxanthine guanine phosphoribosyl transferase 1 (*Hprt*) and represented as the fold induction over naive.

Microarray analysis, REA and LM-PCR. Described in Methods.

ChIP. 10 × 10⁶ BMMΦ were stimulated, washed with PBS, and fixed. Fixed nuclei were sonicated to obtain fragments ranging from 200 to 700 bp. Sonicates were incubated with antibody overnight, followed by Protein A/G beads for 3 h. Recovered DNA was extracted, precipitated, and amplified by qPCR.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 March; accepted 5 April 2007.

Published online 30 May 2007.

1. Nathan, C. Points of control in inflammation. *Nature* 420, 846–852 (2002).

2. Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annu. Rev. Immunol.* **21**, 335–376 (2003).
3. Nieuworp, M., Stroes, E. S., Meijers, J. C. & Buller, H. Hypercoagulability in the metabolic syndrome. *Curr. Opin. Pharmacol.* **5**, 155–159 (2005).
4. Karin, M., Lawrence, T. & Nizet, V. Innate immunity gone awry: Linking microbial infections to chronic inflammation and cancer. *Cell* **124**, 823–835 (2006).
5. Liew, F. Y., Xu, D., Brint, E. K. & O'Neill, L. A. Negative regulation of toll-like receptor-mediated immune responses. *Nature Rev. Immunol.* **5**, 446–458 (2005).
6. Beeson, P. B. Tolerance to bacterial pyrogens I: Factors influencing its development. *J. Exp. Med.* **86**, 29–38 (1947).
7. Beeson, P. B. Tolerance to bacterial pyrogens II: Role of the reticulo-endothelial system. *J. Exp. Med.* **86**, 39–44 (1947).
8. West, M. A. & Heagy, W. Endotoxin tolerance: A review. *Crit. Care Med.* **30**, S64–S73 (2002).
9. Cavaillon, J. M. & Adib-Conquy, M. Bench to bedside review: Endotoxin tolerance as a model of leukocyte reprogramming in sepsis. *Crit. Care* **10**, 1–8 (2006).
10. Dobrovolskaia, M. A. & Vogel, S. N. Toll receptors, CD14, and macrophage activation and deactivation by LPS. *Microbes Infect.* **4**, 903–914 (2002).
11. Fujihara, M. et al. Molecular mechanisms of macrophage activation and deactivation by lipopolysaccharide: roles of the receptor complex. *Pharmacol. Ther.* **100**, 171–194 (2003).
12. Medvedev, A. E., Kopydlowski, K. M. & Vogel, S. N. Inhibition of lipopolysaccharide-induced signal transduction in endotoxin-tolerized mouse macrophages: Dysregulation of cytokine, chemokine, and toll-like receptor 2 and 4 gene expression. *J. Immunol.* **164**, 5564–5574 (2000).
13. Medvedev, A. E., Lentschat, A., Wahl, L. M., Golenbock, D. T. & Vogel, S. N. Dysregulation of LPS-induced Toll-like receptor 4-MyD88 complex formation and IL-1 receptor-associated kinase 1 activation in endotoxin-tolerant cells. *J. Immunol.* **169**, 5209–5216 (2002).
14. Fujihara, M. et al. Lipopolysaccharide-triggered desensitization of TNF- α mRNA expression involves lack of phosphorylation of I κ B α in a murine macrophage-like cell line, P388D1. *J. Leukoc. Biol.* **68**, 267–276 (2000).
15. Dobrovolskaia, M. A. et al. Induction of *in vitro* reprogramming by Toll-like receptor (TLR)2 and TLR4 agonists in murine macrophages: Effects of TLR “homotolerance” versus “heterotolerance” on NF- κ B signaling pathway components. *J. Immunol.* **170**, 508–519 (2003).
16. Huang, Q. et al. The plasticity of dendritic cell responses to pathogens and their components. *Science* **294**, 870–875 (2001).
17. Ogawa, S. et al. Molecular determinants of crosstalk between nuclear receptors and toll-like receptors. *Cell* **122**, 707–721 (2005).
18. De Bosscher, K., Vanden Berghe, W. & Haegeman, G. The interplay between the glucocorticoid receptor and nuclear factor- κ B or activator protein-1: Molecular mechanisms for gene repression. *Endocr. Rev.* **24**, 488–522 (2003).
19. Narlikar, G. J., Fan, H. Y. & Kingston, R. E. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**, 475–487 (2002).
20. Chi, T. A BAF-centred view of the immune system. *Nature Rev. Immunol.* **4**, 965–977 (2004).
21. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
22. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
23. Kurdistani, S. K. & Grunstein, M. Histone acetylation and deacetylation in yeast. *Nature Rev. Mol. Cell Biol.* **4**, 276–284 (2003).
24. Leoni, F. et al. The histone deacetylase inhibitor ITF2357 reduces production of pro-inflammatory cytokines *in vitro* and systemic inflammation *in vivo*. *Mol. Med.* **11**, 1–15 (2005).
25. Nusinzon, I. & Horvath, C. M. Unexpected roles for deacetylation in interferon- and cytokine-induced transcription. *J. Interferon Cytokine Res.* **25**, 745–748 (2005).
26. Santos-Rosa, H. et al. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
27. Schneider, R. et al. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biol.* **6**, 73–77 (2004).
28. Pokholok, D. K. et al. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527 (2005).
29. Bernstein, B. E. et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
30. Metzger, E. et al. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436–439 (2005).
31. Ramirez-Carrozzi, V. R. et al. Selective and antagonistic functions of SWI/SNF and Mi-2 β nucleosome remodeling complexes during an inflammatory response. *Genes Dev.* **20**, 282–296 (2006).
32. Kaufmann, A., Gemsa, D. & Sprenger, H. Differential desensitization of lipopolysaccharide-inducible chemokine gene expression in human monocytes and macrophages. *Eur. J. Immunol.* **30**, 1562–1567 (2000).
33. Learn, C. A., Mizel, S. B. & McCall, C. E. mRNA and protein stability regulate the differential expression of pro- and anti-inflammatory genes in endotoxin-tolerant THP-1 cells. *J. Biol. Chem.* **275**, 12185–12193 (2000).
34. Henricson, B. E., Manthey, C. L., Perera, P. Y., Hamilton, T. A. & Vogel, S. N. Dissociation of lipopolysaccharide (LPS)-inducible gene expression in murine macrophages pretreated with smooth LPS versus monophosphoryl lipid A. *Infect. Immun.* **61**, 2325–2333 (1993).
35. Shnyra, A., Brewington, R., Alipio, A., Amura, C. & Morrison, D. C. Reprogramming of lipopolysaccharide-primed macrophages is controlled by a counterbalanced production of IL-10 and IL-12. *J. Immunol.* **160**, 3729–3736 (1998).
36. Flohé, S. et al. Endotoxin tolerance in rats: Expression of TNF- α , IL-6, IL-10, VCAM-1 and HSP 70 in lung and liver during endotoxin shock. *Cytokine* **11**, 796–804 (1999).
37. Smale, S. T. The establishment and maintenance of lymphocyte identity through gene silencing. *Nature Immunol.* **4**, 607–615 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Smale, T. Chi, M. Wan and R. Rutishauser for discussions, gifts of reagents, and technical assistance. S.L.F. is supported by the UNCF-Merck Graduate Science Research Dissertation Fellowship and by the NIH. D.C.H. is supported by the NSF and the graduate programme at Yale University. R.M. is supported by funding from the Howard Hughes Medical Institute, and the NIH.

Author Information All microarray data are available from the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo) under accession code GSE7348. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.M. (ruslan.medzhitov@yale.edu).

METHODS

Mice. C57BL/6 mice were obtained from Jackson Laboratories. Mice were maintained at the animal facility of Yale University School of Medicine and used at 8–12 weeks of age.

Reagents. LPS, 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB), polymyxin B sulphate (PB), trichostatin A (TSA), dexamethosone (Dex), actinomycin D (ActD), parglyine (Parg) and cycloheximide (CHX) were purchased from Sigma. All restriction enzymes were purchased from New England BioLabs. Dimethyl adipimidate (DMA) was purchased from Pierce. p38 and JNK inhibitors were purchased from Calbiochem. Recombinant mouse IL-6 was purchased from R&D Systems. Paired antibodies for IL-6 and antibodies to phosphorylated p38, ERK, and JNK were purchased from BD Biosciences. Antibodies to H4-Ac (06-866), H3K4me3 (07-473) were purchased from Upstate Biotechnologies. Antibodies to RNA polymerase II (N-20) (sc-899X), p65 (C-20) (sc-372X), and I κ B α were purchased from Santa Cruz Biotechnologies. Antibody to Grp94 was purchased from Stressgen. Antibodies to Brg1 (J1) and Mi-2 β were a kind gift from T. Chi (Yale Univ.) and S.T. Smale (UCLA), respectively.

Bone marrow macrophage (BMM Φ) cultures. Bone marrow progenitors were harvested from mice and cultured for 7 days on Petri dishes in M-CSF supplemented RPMI-1640. Cells were lifted with cold PBS and replated on tissue-culture treated plates. On day 8, macrophages were left untreated (naive, N) or stimulated with 100 ng ml $^{-1}$ LPS for 24 h (tolerant, T), washed twice with warm PBS and given fresh media (N, T) or 10 ng ml $^{-1}$ LPS (N+L, T+L). Where indicated, BMM Φ were treated with Dex (1 μ M), ActD (10 μ g ml $^{-1}$), TSA (50 nM), DRB (15 μ M), Parg (3 μ M), PB (50 μ g ml $^{-1}$) and CHX (100 μ g ml $^{-1}$). The p38/JNK inhibitor cocktail included PD98059 (45 μ M), SB203580 (2.5 μ M), SB202190 (0.5 μ M), PD169136 (5 μ M) and JNK inhibitor II (500 nM).

ELISA. Supernatants were collected 24 h after stimulation and IL-6 was detected with paired antibodies, using recombinant protein to generate a standard curve. Antibody binding was detected by streptavidin-horseradish peroxidase (Zymed) and developed with o-phenylenediamine dihydrochloride (Sigma).

Reverse transcription and quantitative PCR (RT-qPCR). Total RNA from BMM Φ was isolated with RNA-bee reagent (Tel-Test). Total RNA was reverse transcribed with an oligo (dT) primer using Superscript reverse transcriptase III (Gibco BRL). Complementary DNA was analysed in triplicate by qPCR amplification using SYBR Green QPCR Master Mix (Qiagen) on the MX3000P QPCR System (Stratagene). The PCR amplification conditions were: 95 °C (15 min), 45 cycles of 94 °C (30 s), 58 °C (30 s) and 72 °C (1 min). Primer pairs were designed to amplify mRNA-specific fragments and unique products were tested by melt-curve analysis. Data was analysed by comparative quantification using MXPro software with naive values set as calibrator and expression normalized to hypoxanthine guanine phosphoribosyl transferase 1 (*Hprt*). Data are represented as the fold induction over naive (unstimulated).

Microarray analysis. Total RNA was isolated with RNA-bee reagent (Tel-Test) and purified with RNEasy Kit (Qiagen). Two biological replicates were performed for each experimental condition. Sample preparation and hybridization to Affymetrix Mouse Genome 430 2.0 arrays were performed at the Yale W.M. Keck Facility. Briefly, target cDNA generated from each sample was biotinylated, hybridized, and stained as per manufacturer's recommendation using an Affymetrix GeneChip Instrument System. Arrays were scanned on an Affymetrix GeneChip scanner 3000 according to Affymetrix standard protocols (GeneChip Expression Analysis Technical Manual, Affymetrix, 2004). Data was processed using Affymetrix Microarray Suite version 5.0, scaled to a target intensity of 500. Raw and normalized data have been submitted to the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>), accession number GSE7348. Data were further analysed using GeneSpring (Silicogenetics). Analysed expression data are presented in Supplementary Table 1. Probesets present on only one of the 2 arrays for each condition were excluded, as were probesets with a detection call of 'absent' or 'marginal' in all three conditions. The signal intensity for each probeset was averaged over the two arrays, and probesets with a signal intensity ratio of less than 2 in stimulated versus unstimulated macrophages were excluded. Class T genes are defined as genes induced in naive macrophages stimulated with LPS and downregulated more than 3-fold in tolerant macrophages stimulated with LPS (that is, (N+L)/(T+L) > 3). Class NT genes are defined as genes induced in naive macrophages stimulated with LPS and expressed at equal or higher levels in tolerant macrophages stimulated with LPS (that is, (N+L)/(T+L) ≤ 1). Data are displayed as 'fold N+L' (signal N+L/signal N), 'fold T+L' (signal T+L/signal N), 'fold N+L/fold T+L' (class T) and 'fold T+L/Fold N+L' (class NT).

Restriction enzyme accessibility (REA) and ligation mediated (LM)-PCR. Isolated cell nuclei were incubated with limiting amounts of a restriction enzyme (AflII(*il-6*); Sac(*fpr1TSS*); Mse(*fpr15'*)) at 37 °C for 10 min, and then treated with proteinase K overnight ('Experimental' digest). DNA was purified and digested to completion at an upstream site (Nhe(*il-6*); Mse(*fpr1TSS*;

Alu(*fpr15'*)) ('Input' digest). Digested fragments were ligated to complementary annealed linkers and amplified using SYBR Green QPCR Master Mix, a 5' primer derived from the linker, and a 3' gene-specific primer. The touchdown PCR amplification conditions were: 95 °C (15 min), 10 cycles of 94 °C (15 s), 70 °C (30 s) (−1 °C per cycle), 72 °C (2 min), then 18 cycles of 94 °C (15 s), 60 °C (30 s), 72 °C (2 min). This reaction was further amplified for 2–3 cycles with a 32 P-labelled nested 3' gene-specific primer. The PCR products were run out on a 6% polyacrylamide gel and visualized by autoradiography. The ratio of the experimental digest (Exp.) to the input digest (Inp.) reflects the degree of accessibility at the experimental restriction site.

Chromatin immunoprecipitation (ChIP). 10×10^6 BMM Φ were stimulated, washed with PBS, and either fixed with 1% formaldehyde for 5 min at 37 °C (Pol II, AcH4, H3K4me3) or fixed with 25 mM DMA for 90 min and then fixed with 1% formaldehyde for 5 min at room temperature (Brg1, p65, Pol II (Fig. 5c only), AcH4 (Fig. 5d only), H3K4me3 (Fig. 5d, f only)). Formaldehyde fixation was stopped with the addition of 1.25 M glycine. Fixed cells were sonicated for either 3 min of 0.5on/0.5off pulses at setting 1, or 2 min 7 s of 0.5on/0.5off pulses at setting 5, to obtain fragments ranging from 200 to 700 bp in size. Sonicates were diluted 5× and incubated with antibody with rotating overnight. Protein A/G beads (Upstate Biotechnologies) were added for 3 h and collected beads were washed extensively. Protein–DNA complexes were eluted from the beads and treated with 200 mM NaCl to reverse cross-links and proteinase K to digest proteins. Recovered DNA was phenol:chloroform extracted and precipitated with isopropanol and glycogen. Immunoprecipitated DNA and input DNA were amplified with gene-specific and β -globin (*Hbb*) primers by qPCR, using input DNA to generate a standard curve. ChIP data is represented as %input(gene-specific)/%input(β -globin), except in Fig. 1e, where it is represented as %input.

LETTERS

Deposition of metal films on an ionic liquid as a basis for a lunar telescope

Ermanno F. Borra¹, Omar Seddiki¹, Roger Angel², Daniel Eisenstein², Paul Hickson³, Kenneth R. Seddon⁴ & Simon P. Worden⁵

An optical/infrared telescope of 20–100 m aperture located on the Moon would be able to observe objects 100 to 1,000 times fainter than the proposed next generation of space telescopes¹. The infrared region of the spectrum is particularly important for observations of objects at redshifts $z > 7$. The apparent simplicity and low mass of a liquid mirror telescope, compared with a traditional pointable glass mirror, suggest that the concept should be considered further. A previously proposed liquid mirror telescope, based upon a spinning liquid metallic alloy², is not appropriate for infrared applications, which will require a liquid below 130 K. Here we report the successful coating of an ionic liquid with silver. The surface is smooth and the silver coating is stable on a timescale of months. The underlying ionic liquid does not evaporate in a vacuum and remains liquid down to a temperature of 175 K. Given that there are $\sim 10^6$ simple and $\sim 10^{18}$ ternary ionic liquids, it should be possible to synthesize liquids with even lower melting temperatures.

The Lunar Liquid Mirror Telescope (LLMT) concept is aimed at furthering our understanding of the early Universe. This has recently been revolutionized by deep optical fields imaged with the Hubble Space Telescope and the Spitzer infrared telescope. In the next two decades, Hubble's successor, the 6-m infrared James Webb Space Telescope¹ (JWST), will extend this work. However, to learn more about much deeper fields, a very cold space telescope will be required, with apertures 20–100 m in diameter and capable of integrations of days to years with Hubble quality imaging.

An LLMT has advantages and disadvantages. Liquid surfaces have the intrinsic advantage that they are smooth if undisturbed and that two fundamental natural forces (gravity and inertia) conspire to restore them to the desired parabolic shape after an external perturbation. Liquid mirrors have excellent optical qualities^{3,4}. Liquid mirror telescopes are simple instruments, so shipping and assembling will be easier than for a solid mirror; there may be other advantages, but a detailed engineering analysis still needs to be done. The Supplementary Information section discusses technical considerations.

The main disadvantage of an LLMT comes from the fact that the field of regard of a zenith telescope is small, limiting the observations it can make. Fortunately, this is a minor problem for selected fields of astronomical research, like cosmology. The Supplementary Information discusses technologies that promise to extend the field of regard. There may also be unforeseen lunar environmental factors affecting the remote coating of the liquid or its longevity. Such problems can be identified and hopefully solved by sending smaller precursor mirrors to the moon.

Primary mirrors made by spinning a liquid in a gravitational field represent a new technology, with the potential to achieve very large

size and high optical quality at low cost. The technology is young, but its performance is well documented by laboratory tests^{3,4} and observations^{5–8}. Zenith-pointing telescope mirrors of liquid mercury have been made at very low cost up to diameters of 6 m. Indeed, it appears that liquid mirror telescopes such as the 6-m-diameter Large Zenith Telescope⁹ (see also www.astro.ubc.ca/lmt/lzt) can be constructed for a few per cent of the cost of corresponding conventional telescope designs of a similar aperture^{10,11}.

Because an LLMT would survey a generic patch of sky, the most interesting applications are for high-redshift targets, where (by definition) there are no closer examples elsewhere in the sky that can be studied with a smaller telescope like JWST. However, as has been well argued for JWST¹, the window between 0.5 and 5 μm is a superb place to observe the high-redshift Universe, as one can track the formation and aggregation of stars as well as the build-up of metallicity through rest-frame optical and ultraviolet signatures. Moreover, with multi-band coverage in the rest-frame ultraviolet band, the Lyman break gives a robust redshift signature¹².

Because of the large aperture, small beam size, and long exposure times, an LLMT could reach point-source flux sensitivities in the 1–10 pJy range for near-infrared broad-band imaging, a factor of 100–1,000 fainter than JWST. This sensitivity and angular resolution would be very useful for dissecting normal and dwarf galaxies at redshifts 1 to 10. However, the most interesting frontier is probably at redshifts 15–20, where the first stars form and the first galaxies assemble¹³. Here, JWST will detect proto-galaxies down to stellar masses of about 10^7 solar masses, but an LLMT could decrease that by a factor of 100–1,000. A 100-m LLMT may even be able to detect the ultraviolet radiation from the photosphere of a single Population III star with a mass of several hundred solar masses^{14,15}. In addition, an LLMT could produce spectroscopy of the stellar continua of proto-galaxies at and below the JWST detection limit, allowing us to probe the ages and metallicities of these early stars as well as the intervening intergalactic medium via metal-line absorption systems.

The LLMT needs a liquid having very low vapour pressure, low melting temperature and high reflectivity. We considered three main classes of liquids that could be used: eutectics of Group 1 metals, which have the advantage of high intrinsic reflectivity², low-reflectivity liquids coated with self-assembling nanoparticles^{16,17}, and low-reflectivity liquids coated with a thin metallic coating produced by sputtering or evaporation. We concluded that the most promising approach for infrared observations was to apply metallic coatings to low-temperature, low-vapour-pressure, moderate-viscosity liquids.

Borrowing from techniques used to coat solid mirrors on Earth, we conducted a number of experiments by vaporization under vacuum

¹Département de Physique, Génie Physique et Optique, Centre d'Optique, Photonique et Lasers, Université Laval, Québec, G1K 7P4, Canada. ²Steward Observatory, The University of Arizona, 933 N. Cherry Avenue, Tucson, Arizona 8575, USA. ³Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada. ⁴The QUILL Centre, The Queen's University of Belfast, Stranmillis Road, Belfast BT9 5AG, UK. ⁵Office of the Director, NASA Ames Research Center, Moffett Field, California 9403, USA.

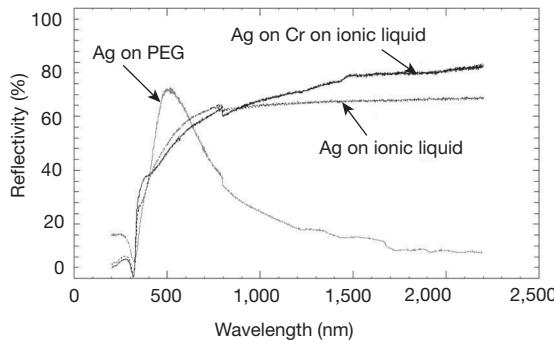


Figure 1 | Reflectivity curves obtained for silver-coated liquids. We conducted a number of coating experiments by vaporizing in vacuum a reflective silver layer on to liquids. The figure shows our best reflectivity curve with a hydrophilic block polymer (PEG). We then successfully coated an ionic liquid, an important milestone because ionic liquids have negligible vapour pressures. The figure shows a reflectivity curve obtained for a silver-coated ionic liquid and for silver deposited on chromium deposited on an ionic liquid. The silver on chromium on ionic-liquid coating is a noticeable improvement over the silver coating on PEG. The curves do not extend beyond 2.2 μm because that is the sensitivity limit of our equipment. Presumably one can extrapolate that the reflectivity increases further in the infrared. We also noted that as the temperature of deposition was lowered, both the quality of the film and its infrared reflectivity improved.

of a reflective metal on liquids. We experimented mostly with silver, which possesses the high reflectivity needed for an LLMT. Before identifying cryogenic liquids suitable for an LLMT, one must determine the physical characteristics needed for successful coating. To our knowledge, liquids had never been vacuum coated before, so we sampled the parameter space using a variety of liquids. Most experiments were unsuccessful, failing to coat the liquids with a reflective layer, but we did succeed in coating a silicone oil (although it gave an unusable wrinkled skin). Then we performed two important experiments. First, building on our experience with nanoparticles^{16,17}, we successfully coated the hydrophilic block copolymer PPG–PEG–PPG (50% PEG; PEG, polyethylene glycol; PPG, polypropylene glycol) with silver. The observed reflectivity curve is shown in Fig. 1 and while being clearly inadequate for our purpose, it was a significant improvement on the coatings on silicone oil.

From the abovementioned coating and related results, it was clear that the ideal liquid to form the base for a highly reflective, uniform metal film would have effectively zero vapour pressure, high viscosity and a low melting point, while being effectively involatile *in vacuo*. This property set is found for some members of a suite of fluids known as ionic liquids^{18,19}. Ionic liquids are salts which are liquid at temperatures below 373 K, composed entirely of ions, and usually possess no significant vapour pressure at room temperature or below^{20–22}. They are also often highly viscous, and may be either hydrophilic or hydrophobic. Following this logic, we successfully coated a hydrophilic commercially-available ionic liquid, 1-ethyl-3-methylimidazolium ethylsulphate ([emim][EtOSO₃]; commercially known as ECOENG 212), which solidifies at 175 K (ref. 23). This was a transformational breakthrough: Fig. 1 shows a reflectivity curve obtained for silver-coated [emim][EtOSO₃]. There is substantial improvement with respect to the curve obtained using PEG-derived liquids as a substrate. Moreover, as there are at least a million (10^6) simple ionic liquids, and a trillion (10^{18}) ternary ionic liquid systems²⁴, there is a phenomenally wide choice for optimizing the properties of the liquid substrate, to minimize melting point and volatility, while obtaining optimal infrared reflectivity.

A major problem that we encountered with the direct deposition of silver on an ionic liquid is that it tends to diffuse in the liquid substrate, making it difficult to obtain a thick layer. We note, however, that this appears to be a problem only during deposition. After deposition, the surface coating is stable. Indeed, electron microscopy

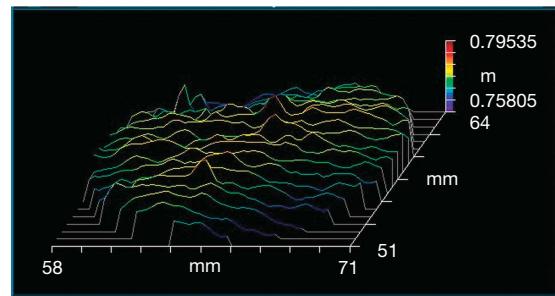


Figure 2 | Three-dimensional map of a small section of a silver-coated liquid mirror. The 1.25 cm² area is made of a 30-nm-thick silver layer deposited on a 5-nm-thick chromium layer deposited on an ionic liquid. It shows a respectfully small peak-to-valley deviation of only 0.0373 μm, which is a hundredth of a wave at 4 μm. This is close to the standard deviation of the 0.03 μm error of measurement of our interferometer. The measurements were made with a Mach–Zender interferometer operating at a wavelength of 632.8 nm.

showed that the film is made of colloidal particles having diameters of a few tens of nanometres. Indeed, ionic liquids have been reported to induce stable colloid formation^{25–27}. This appears to be the reason why the reflectivities seen in Fig. 1 are lower than the reflectivity of metallic silver. Our next successful attempt to improve film reflectivity involved the initial deposition of a chromium film, followed by subsequent deposition of a silver layer on the chromium layer. The nucleation density is greater for chromium than it is for silver, so chromium is far easier to deposit than silver. The thickness of a chromium layer deposited on an ionic liquid grows significantly faster than an analogous silver layer, and a silver layer deposited on this intermediate chromium layer grows substantially faster than a silver layer directly deposited on an ionic liquid. Figure 1 shows the reflectivity curve of a silver-on-chromium-on-ionic-liquid mirror, which is significantly better than the reflectivity curve of a silver-on-ionic-liquid mirror. Figure 2 shows the three-dimensional map of a small (1.25 cm²) section of a mirror made of a 30-nm-thick silver layer deposited on a 5 nm chromium layer deposited on the [emim][EtOSO₃] ionic liquid. The optical quality of the surface is excellent.

Although the reflectivities shown in Fig. 1 are not yet adequate, it is now only a matter of technological improvement. This will require better vacuum facilities and more experiments to improve the coating technique and reflectivity. We have already found that the reflectivities shown in Fig. 1 can be improved by increasing the thickness of the silver layer. However, heating of the metallic layer from the heating element prevents us from increasing the thickness at present. We are thus experimenting with cooling techniques, but for technical reasons, they are difficult to implement in a vacuum tank. We will also be experimenting with different coating techniques.

We have shown here that various liquids can be successfully coated with a reflective metal surface. Moreover, for the coated liquid to function as a liquid primary telescope mirror on the Earth's moon, it is essential that the liquid have a low vapour pressure as well as a low freezing temperature. These two requirements have been met by using an ionic liquid. In fact, the metal coating of an ionic liquid has been a defining demonstration, since it is essential for the potential implementation of an infrared LLMT—an instrument which will revolutionize astronomical observations of the early Universe.

Received 15 November 2006; accepted 4 May 2007.

- Gardner, J. P. et al. The James Webb Space Telescope. *Space Sci. Rev.* 123, 485–606 (2006).
- Borra, E. F. The case for liquid mirror in a lunar telescope. *Astrophys. J.* 373, 317–321 (1991).
- Girard, L. & Borra, E. F. Optical tests of a 2.5-m diameter liquid mirror. II. Behavior under external perturbations and scattered light measurements. *Appl. Opt.* 36, 6278–6288 (1997).

4. Ninane, N. M., & Jamar, C. A. Parabolic liquid mirrors in optical shop testing. *Appl. Opt.* **35**, 6131–6139 (1996).
5. Hickson, P. & Mulrooney, M. K. University of British Columbia—NASA multi-narrowband survey. I. Description and photometric properties of the survey. *Astrophys. J.* **115** (Suppl.), 35–42 (1998).
6. Cabanac, R. Borra, E. F. & Beauchemin, M. A search for peculiar objects with the NASA Orbital Debris Observatory 3-m Liquid Mirror Telescope. *Astrophys. J.* **509**, 309–323 (1998).
7. Hickson, P. Applied optics, hydrodynamics of rotating liquid mirrors. I. Synchronous disturbances. *Appl. Opt.* **45**, 8052–8062 (2006).
8. Hickson, P. & Racine, R. Image quality of liquid-mirror telescopes. *Publ. Astron. Soc. Pacif.* **119**, 456–465 (2007).
9. Hickson, P. *et al.* The Large Zenith Telescope—a 6-meter liquid-mirror telescope. *Publ. Astron. Soc. Pacif.* **119**, 444–455 (2007).
10. Hickson, P. & Lanzetta, K. M. Large Aperture Mirror Array (LAMA): conceptual design for a distributed-aperture 42-meter telescope. *Proc. SPIE* **4840**, 273–282 (2003).
11. Hickson, P. & Lanzetta, K. M. Large Aperture Mirror Array (LAMA): project overview. *Proc. SPIE* **532**, 115–125 (2004).
12. Giavalisco, M. M. Lyman-break galaxies. *Annu. Rev. Astron. Astrophys.* **40**, 579–642 (2002).
13. Barkana, R. & Loeb, A. In the beginning: the first sources of light and the reionization of the universe. *Phys. Rep.* **349**, 125–238 (2001).
14. Bromm, V., Coppi, P. S. & Larson, R. B. Forming the first stars in the universe: the fragmentation of primordial gas. *Astrophys. J.* **527**, L5–L8 (1999).
15. Abel, T., Bryan, G. L. & Norman, M. L. The formation and fragmentation of primordial molecular clouds. *Astrophys. J.* **540**, 39–44 (2000).
16. Borra, E. F. *et al.* Nanoengineered astronomical optics. *Astron. Astrophys.* **419**, 777–782 (2004).
17. Déry, J.-P., Gingras, J., Yockell-Lelièvre, H., Borra, E. F. & Ritcey, A. M. Characterization of reflective silver nanoparticle surface films. *Colloids Surf. A* **279**, 79–86 (2006).
18. Wasserscheid, P. & Welton, T. (eds) *Ionic Liquids in Synthesis* (Wiley-VCH, Weinheim, 2003).
19. Stark, A. & Seddon, K. R. in *Kirk-Othmer Encyclopaedia of Chemical Technology* (ed. Seidel, A.) 836–920 (John Wiley & Sons, Hoboken, New Jersey, 2007).
20. Earle, M. J. *et al.* The distillation and volatility of ionic liquids. *Nature* **439**, 831–834 (2006).
21. Deetlefs, M. & Seddon, K. R. Ionic liquids: fact and fiction. *Chim. Oggi* **24**, 16–23 (2006).
22. MacFarlane, D. R. & Seddon, K. R. Ionic liquids—Progress on the fundamental issues. *Aust. J. Chem.* **60**, 3–5 (2007).
23. Holbrey, J. D. *et al.* Efficient, halide free synthesis of new, low cost ionic liquids: 1,3-dialkylimidazolium salts containing methyl- and ethyl-sulfate anions. *Green Chem.* **4**, 407–413 (2002).
24. Seddon, K. R. in *The International George Papathodorou Symposium: Proceedings* (eds Boghosian, S., Dracopoulos, V., Kontoyannis, C. G. & Voyatzis, G. A.) 131–135 (Institute of Chemical Engineering and High Temperature Chemical Processes, Patras, 1999).
25. Fonseca, G. S. *et al.* Synthesis and characterization of catalytic iridium nanoparticles in imidazolium ionic liquids. *J. Colloid Interf. Sci.* **301**, 193–204 (2006).
26. Itoh, H., Naka, K. & Chujo, Y. Synthesis of gold nanoparticles modified with ionic liquid based on the imidazolium cation. *J. Am. Chem. Soc.* **126**, 3026–3027 (2004).
27. Warren, S. C. *et al.* Generalized route to metal nanoparticles with liquid behavior. *J. Am. Chem. Soc.* **128**, 12074–12075 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank NASA for a NIAC grant and the Canadian Space Agency for primary funding. We also gratefully acknowledge the input from M. Deetlefs; K.R.S. also thanks the EPSRC for support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.F.B. (borra@phy.ulaval.ca).

LETTERS

Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates

Yuriy Román-Leshkov¹, Christopher J. Barrett¹, Zhen Y. Liu¹ & James A. Dumesic¹

Diminishing fossil fuel reserves and growing concerns about global warming indicate that sustainable sources of energy are needed in the near future. For fuels to be useful in the transportation sector, they must have specific physical properties that allow for efficient distribution, storage and combustion; these properties are currently fulfilled by non-renewable petroleum-derived liquid fuels. Ethanol, the only renewable liquid fuel currently produced in large quantities, suffers from several limitations, including low energy density, high volatility, and contamination by the absorption of water from the atmosphere. Here we present a catalytic strategy for the production of 2,5-dimethylfuran from fructose (a carbohydrate obtained directly from biomass or by the isomerization of glucose) for use as a liquid transportation fuel. Compared to ethanol, 2,5-dimethylfuran has a higher energy density (by 40 per cent), a higher boiling point (by 20 K), and is not soluble in water. This catalytic strategy creates a route for transforming abundant renewable biomass resources^{1,2} into a liquid fuel suitable for the transportation sector, and may diminish our reliance on petroleum.

The rationale for converting carbohydrates to 2,5-dimethylfuran (DMF) is outlined in Fig. 1. The selective removal of five oxygen

atoms from a hexose (for example, fructose) to produce DMF not only decreases the boiling point to a value suitable for liquid fuels, but also attains the lowest water solubility and the highest research octane number³ (RON) of the mono-oxygenated C₆ compounds, while preserving a high energy density (30 kJ cm⁻³). This selective removal of oxygen atoms can be accomplished in two steps: first, removing three oxygen atoms by dehydration to produce 5-hydroxymethylfurfural (HMF); and second, removing two oxygen atoms by hydrogenolysis to produce DMF by way of intermediates **4** and **5**. Species **6**, produced by way of **7**, is a hydrogenolysis by-product that also possesses excellent fuel qualities.

A schematic representation of our process is outlined in Fig. 2. (An annotated version of this figure, summarizing the primary findings of this work, is presented in Supplementary Information.) The first step involves the acid-catalysed dehydration of fructose to produce HMF in a biphasic reactor. Because the normal boiling point of HMF is too high for it to be used as a fuel (Fig. 1), the HMF extracted by the organic phase of the biphasic reactor is subsequently converted to DMF by hydrogenolysis of C–O bonds over a copper-ruthenium (CuRu) catalyst. As described below, our process also involves two separation steps.

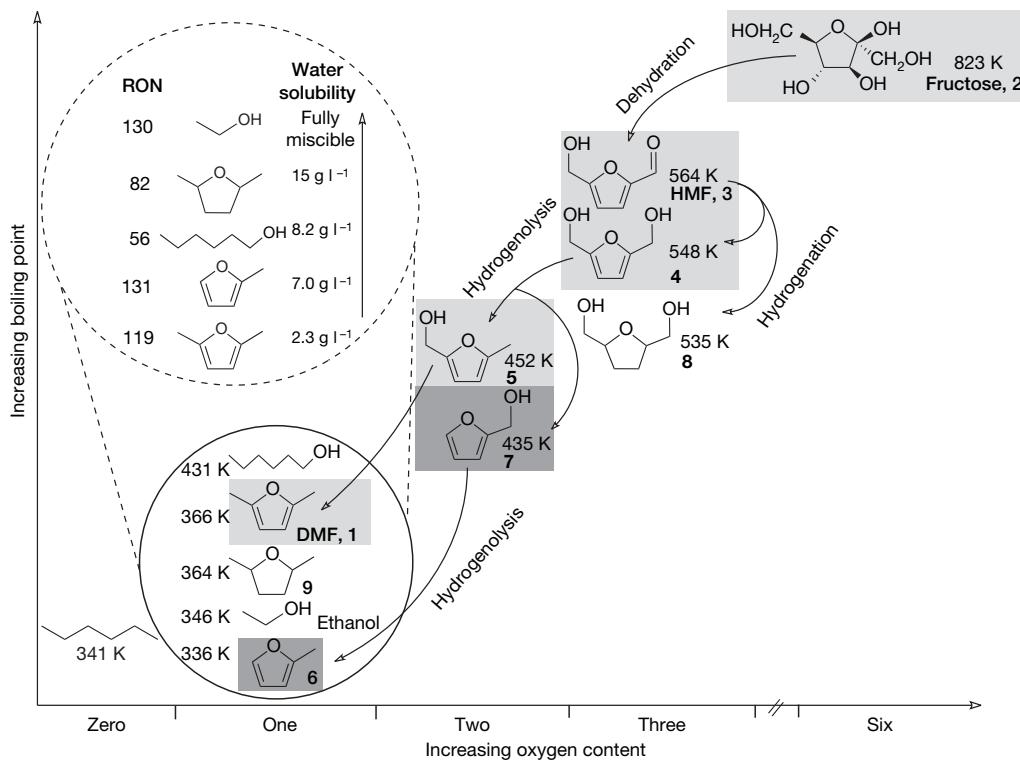
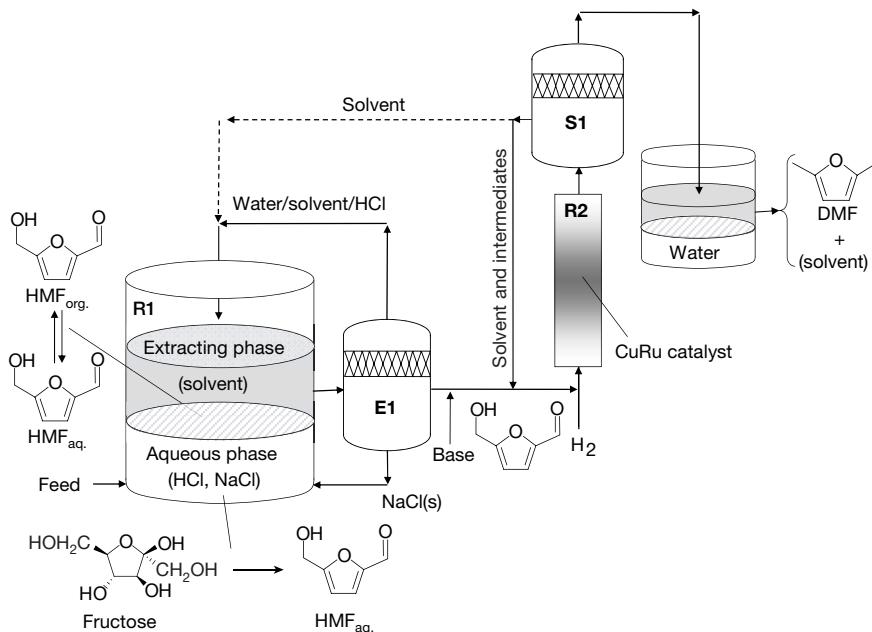


Figure 1 | Normal boiling points of representative C₆-hydrocarbons formed by removal of oxygen atoms from hexoses, compared to the normal boiling point of ethanol. Compounds as follows: 2,5-dimethylfuran (DMF, **1**); D-fructose (**2**); 5-hydroxymethylfurfural (HMF, **3**); 2,5-dihydroxymethylfuran (**4**); 2-methyl,5-hydroxymethylfuran (**5**); 2-methylfuran (**6**); furfural alcohol (**7**); 2,5-dihydroxymethyltetrahydrofuran (**8**); and 2,5-dimethyltetrahydrofuran (**9**). Light grey boxes highlight reactant, product and primary intermediates. Dark grey boxes highlight key by-products and by-product intermediates. Inset, water solubility and research octane number (RON)^{3,24,25} of mono-oxygenated C₆-compounds and **6** compared to ethanol.

¹Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.



To implement our process, we developed a new catalytic system to produce HMF in high yields from concentrated sugar solutions (R1 in Fig. 2). Although multiple catalytic systems have been reported for the production of HMF in solvents containing high boiling point components (for example, dimethylsulphoxide) to suppress side reactions^{4–9}, trace amounts of such components are detrimental to fuel quality, and their removal necessitates energy-intensive purification procedures. We show that HMF can be produced in high yields by the acid-catalysed dehydration of fructose in a biphasic reactor using low boiling point solvents that themselves are excellent fuel components, thereby eliminating the need for expensive separation steps to produce the final liquid fuel mixture. The reactive aqueous phase in the biphasic reactor contains an acid catalyst and a sugar, and the extracting phase contains a partially miscible organic solvent (for example, butanol) that continuously extracts the HMF product. Importantly, the addition of a salt to the aqueous phase improves the partitioning of HMF into the extracting phase, and leads to increased HMF yields without the use of high boiling point solvents.

The capacity of the organic phase to extract HMF from the reactive aqueous phase, as measured by the extracting ratio R (the ratio of the HMF concentration in the organic layer to that in the aqueous layer), directly affects HMF selectivity, defined as the moles of HMF produced divided by the moles of fructose reacted (Fig. 3, Table 1). Our results demonstrate that HMF selectivity increases as R increases, indicating that a more efficient removal of HMF from the aqueous phase prevents undesired side reactions. The addition of salt to the reactive aqueous phase increases R by means of the salting-out effect, whereby electrolytes alter the intermolecular bonding interactions between liquid components, decreasing the mutual solubility of the aqueous and organic phases^{10,11}. The R value for a specific extracting solvent depends not only on the affinity of the solvent for HMF, but also on the ability of the salt to separate both phases. In the present work, the addition of NaCl to the aqueous phase resulted in the largest increase in R of all the salts tested; results for other salts can be found in Supplementary Information. For example, compared to experiments without salt, a 30 wt% fructose solution saturated with NaCl (35 g of NaCl per 100 g of H₂O) using 2-butanol as the extracting solvent (with initial ratio of organic and aqueous phase volumes $V_{\text{org}}/V_{\text{aq}} = 1.6$) results in an increase in R from 1.6 to 3.3, leading to an improvement in HMF selectivity from 66% to 79% (Table 1, runs 1 and 6). Notably, the presence of NaCl has the additional benefit of allowing higher values of $V_{\text{org}}/V_{\text{aq}}$ to be used, thus leading to higher HMF selectivities, while maintaining biphasic reaction conditions.

Figure 2 | Schematic diagram of the process for conversion of fructose to DMF. Diagram includes selective dehydration of fructose to form HMF in a biphasic reactor (R1); evaporation of water and HCl from the liquid solvent containing HMF, leading to precipitation of NaCl (E1); hydrogenolysis of HMF to DMF over a CuRu catalyst (R2); and separation of the DMF product from the extracting solvent and unreacted intermediates (S1). See text for details.

Specifically, when the ratio $V_{\text{org}}/V_{\text{aq}}$ is doubled, the 2-butanol system without salt becomes monophasic, whereas the system saturated with NaCl remains biphasic, with an R of 3.6 and an HMF selectivity of 89% (Table 1, run 5). The primary role of NaCl is to alter the solvent properties while remaining otherwise inert. Specifically, the dehydration of fructose in the presence of NaCl, but in the absence of an extracting solvent, leads to the same HMF selectivity as in the absence of NaCl (Table 1, runs 19 and 20).

Although various extracting solvents can generate high HMF selectivity (Fig. 3), the use of 1-butanol as a solvent is advantageous for biomass applications. For instance, solvents such as 2-butanol are obtained from petroleum-derived products (by the hydrolysis of 2-butene; ref. 12), whereas 1-butanol can be produced by the fermentation of biomass-derived carbohydrates^{13,14}. Also, unlike unsaturated solvents such as toluene or methylisobutylketone, 1-butanol is inert in the hydrogenolysis step of our process. The dehydration

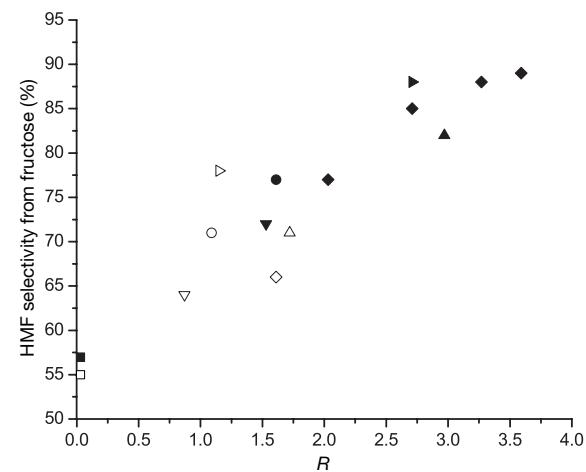


Figure 3 | Effect of extraction ratio R on HMF selectivity from fructose for various organic solvents. Open symbols, experiments without NaCl; filled symbols, experiments with an aqueous phase saturated with NaCl. Diamonds, 2-butanol (filled diamonds, experiments using 2-butanol as the extracting solvent and aqueous phases containing 5, 15, 25 and 35% NaCl; open diamond, experiments using 2-butanol with no salt and a $V_{\text{org}}/V_{\text{aq}} = 1.6$); upward-pointing triangles, 1-butanol; downward-pointing triangles, 1-hexanol; circles, methylisobutylketone; right-pointing triangles, 5:5 toluene:2-butanol; squares, no solvent. $R = [\text{HMF}]_{\text{org}}/[\text{HMF}]_{\text{aq}}$.

reaction using 1-butanol and an aqueous phase saturated with NaCl shows an *R* value of 3.0 and an HMF selectivity of 82% (Table 1, run 8). When using 1-butanol, we observed that HMF selectivity was independent of acid content for experiments in the range of HCl concentrations from 0.01 to 0.25 M (Table 1, runs 8–12).

The extracting solvent containing HMF next undergoes a purification step (E1 in Fig. 2). For an experiment analogous to run 12 in Table 1, the stream entering the evaporator contains 260 mmol l⁻¹ HMF (as measured by high performance liquid chromatography, HPLC), 2,800 mmol l⁻¹ water (see Supplementary Information for calculation assumptions), 26 mmol l⁻¹ NaCl (as measured by HPLC), 3.6 mmol l⁻¹ HCl (as measured by acid titration with NaOH), and 1.1 mmol l⁻¹ 1-chlorobutane (as measured by gas chromatography). Vacuum evaporation at low temperature (for example, 363 K) can be used to separate volatile impurities from components with lower relative volatility, such as HMF and 1-butanol. Furthermore, because the solubility of NaCl in anhydrous organic solvents is lower than in water, removal of water causes the salt to precipitate out of solution. Evaporation of 25% of the total mass of the entering stream lowers the impurity levels to 360 mmol l⁻¹ water, 1.6 mmol l⁻¹ NaCl, 1.5 mmol l⁻¹ HCl, and 0.2 mmol l⁻¹ 1-chlorobutane. Levels of 1-chlorobutane are lowered below our detection limits (<0.1 mmol l⁻¹) by neutralization with NaOH of the purified stream leaving the evaporator. Thus, the water, the NaCl, the fraction of 1-butanol that evaporates, and 58% of the HCl are recovered and recycled back into the biphasic reactor, whereas the purified liquid stream containing HMF and 1-butanol is sent for further processing.

Next, HMF is converted to DMF over a copper-based catalyst (R2 in Fig. 2). Previous studies have shown that copper chromite (CuCrO₄) selectively converts furfural into 7 and 6 without excessive hydrogenation of the furan ring or excessive ring decomposition products^{15–17}. Accordingly, CuCrO₄ should be an effective catalyst for the hydrogenolysis of HMF to DMF, although no studies of this reaction have been reported. Our liquid-phase batch experiments of

HMF hydrogenolysis using CuCrO₄ showed 61% yield (defined as the product of selectivity and conversion) for DMF and 29% yield for 5 (detailed results can be found in Supplementary Information). Importantly, however, trace levels of chloride ions in the solvent (introduced during the dehydration step and not completely removed during the evaporation step) deactivate the CuCrO₄ catalyst significantly. For instance, when this catalyst is used in a 1-butanol solution containing 1.6 mmol l⁻¹ of NaCl, only a 6% yield of DMF is obtained. The literature regarding the deactivation of copper-based catalysts by p.p.m. levels of chloride species indicates that the primary mode of deactivation is chloride-induced sintering of copper¹⁸.

To alleviate poisoning of the copper catalyst, we developed a chloride-resistant carbon-supported copper-ruthenium (CuRu/C) catalyst. The rationale for using this catalyst was that we first observed that a carbon-supported ruthenium catalyst was resistant to deactivation in the presence of chloride ions; however, this catalyst converted HMF primarily to 8. Because copper and ruthenium are immiscible, and copper has a lower surface energy than ruthenium, their mixture creates a two-phase system in which the copper phase coats the surface of the ruthenium phase^{19–21}. Accordingly, we hypothesized that a CuRu/C catalyst may exhibit copper-like hydrogenolysis behaviour combined with ruthenium-like chlorine resistance.

Liquid-phase hydrogenolysis experiments using a 3:1 (atomic ratio) Cu:Ru/C catalyst produce yields of 71% DMF, 4% 6, and 12% intermediates. Notably, the same catalyst used with a purified 1-butanol solution containing 1.6 mmol l⁻¹ of NaCl generates yields of 61% DMF, 4% 6, and 20% intermediates. Thus, although CuRu/C is affected to some extent by the presence of chloride species, its performance is markedly superior to that of CuCrO₄.

Alternatively, because NaCl does not evaporate, vapour-phase hydrogenolysis experiments were performed using a flow reactor to eliminate effects of chloride ions on CuRu/C. Vapour-phase hydrogenolysis using a 3:2 Cu:Ru/C catalyst shows yields of 76–79% DMF and ~5% intermediates for 1.5 and 10 wt% HMF feeds (see

Table 1 | Dehydration results for 30 wt% fructose solutions

Run	Salt (%)	Organic phase	Conversion (%)	Selectivity (%)	[HMF] _{aq.} (g l ⁻¹)	[HMF] _{org.} (g l ⁻¹)	<i>R</i>	[Salt] _{org.} (g l ⁻¹)	[H ₂ O] _{org.} (wt%)
1	0*	2-butanol	58	66	28.6	46.0	1.6	0.0	31.4
2	5	2-butanol	65	77	16.8	34.1	2.0	0.9	16.4
3	15	2-butanol	65	85	12.7	34.4	2.7	1.1	9.6
4	25	2-butanol	75	88	11.6	37.9	3.3	1.2	6.8
5	35	2-butanol	74	89§	10.6	38.1	3.6	1.6	6.5
6	35*	2-butanol	71	79	18.0	60.0	3.3	1.6	7.4
7	0	1-butanol	52	71	15.1	26.0	1.7	0.0	23.1
8	35	1-butanol	85	82§	13.2	39.2	3.0	1.6	6.1
9	35†	1-butanol	80	83	12.0	39.0	3.3	1.6	6.1
10	35†	1-butanol	88	82	12.9	43.1	3.3	1.6	6.1
11	35†	1-butanol	77	84	12.4	37.8	3.0	1.6	6.1
12	35†	1-butanol	64	84	10.2	32.4	3.2	1.6	6.1
13	0	1-hexanol	50	64	21.1	18.4	0.9	0.0	7.9
14	35	1-hexanol	78	72	19.5	29.9	1.5	0.9	2.2
15	0	MIBK	50	71	20.0	21.8	1.1	0.0	0.9
16	35	MIBK	72	77	18.3	29.3	1.6	0.2	0.0
17	0	Toluene:2-butanol	64	78	27.7	31.7	1.2	0	6.7
18	35	Toluene:2-butanol	74	88	13.8	37.4	2.7	0.8	1.9
19	0	None	44	55	53.5	0.0	0.0	0.0	-
20	35	None	59	57	70.8	0.0	0.0	35.0	-
21	5‡	2-butanol	30	36	1.2	2.3	1.9	0.9	16.4
22	35‡	2-butanol	56	48	1.1	3.9	3.6	1.6	6.5

Fructose wt% is calculated on a salt-free basis. Standard reaction conditions: *T* = 453 K and *V*_{org.}/*V*_{aq.} = 3.2 with 0.25 M HCl catalyst (mol HCl per l of aqueous phase). Salt % is expressed as grams of salt divided by grams of water × 100.

* Runs that used *V*_{org.}/*V*_{aq.} = 1.6.

† Runs 9–12 used 0.12, 0.06, 0.03 and 0.01 M HCl, respectively.

‡ Runs that used a 10 wt% glucose (salt-free basis) feed.

§ Selectivity values for 1-butanol and 2-butanol systems saturated with NaCl are reported as means with s.d. of ±1.3% and ±1.5%, respectively (*n* = 5).

|| 5:5 mass ratio.

Supplementary Information for detailed results). No chlorinated hydrocarbons were detected after reaction. Thus, although the vapour-phase process requires vaporization of the feed, it offers multiple benefits. First, when compared to the liquid-phase process, it generates no by-products and fewer intermediates. Second, it can process both dilute and concentrated HMF solutions. Third, because the same yields were obtained when using 1-butanol or 1-hexanol, other solvents can be used without altering the selectivity. Last, although the catalyst slowly deactivates after processing an amount of HMF equivalent to 1.7 times the mass of the catalyst, it can be regenerated fully by flowing hydrogen at the reaction temperature.

DMF can optionally be hydrogenated to **9** over a ruthenium catalyst. **9** contains a higher hydrogen to carbon ratio than DMF, which translates into a higher energy content. Moreover, **9** may provide additional stability on storage over extended periods of time because it contains a fully hydrogenated furan ring. The toxicological properties of neither DMF nor **9** have been thoroughly tested. The limited information available suggests that DMF is not more toxic than current fuel components; however, long-term studies must be performed before these furan components are approved for commercial use (see Supplementary Information for details).

The final step involves the separation of DMF from the solvent and the reaction intermediates (S1 in Fig. 2). The more volatile components (that is DMF, **6** and water) can be separated from the solvent and the intermediates; the latter stream can be then recycled back to the hydrogenolysis reactor. On condensation, the hydrophobic products DMF and **6** separate spontaneously from water. Depending on the final fuel composition requirements, a distillation process may be used to control more precisely the distribution of components and also to recycle a fraction of the solvent to the dehydration reactor. We note that the energy required to evaporate the stream containing DMF and 1-butanol, leading to product separation, is approximately one-third of the energy required to evaporate an aqueous solution of ethanol produced by fermentation for biofuel applications (see Supplementary Information).

The efficiency of wide-scale production of DMF from biomass would improve if a more readily available feedstock, such as glucose, was used. Although we have achieved moderate yields of HMF directly from glucose (Table 1, run 22), the conversion of fructose is more selective. In this respect, efficient processes already exist to produce high levels of fructose from glucose, such as enzyme-catalysed isomerization combined with a simulated moving-bed separation^{22,23}. Although challenges remain for commercial application, this research opens a new path for the production of DMF as a biomass-derived liquid transportation fuel.

METHODS SUMMARY

Dehydration experiments. In a typical experiment, a hexose-containing aqueous solution (for example, 30 wt% fructose on a salt-free basis), an inorganic salt (for example, NaCl), an acid catalyst (for example, 0.25 M HCl) and an organic solvent (for example, 1-butanol) were mixed in a thick-walled glass reactor. The reaction was carried out at 453 K, at the total pressure generated by the vapour pressures of the volatile components at this temperature, and a constant initial volumetric ratio of organic and aqueous phases ($V_{\text{org}}/V_{\text{aq}}$). A reaction time of approximately 3 min was used to convert ~75% of the feed.

Hydrogenolysis experiments. Liquid-phase hydrogenolysis experiments were carried out using 5 wt% HMF in a 1-butanol solution at 493 K and 6.8 bar H_2 pressure. Vapour-phase experiments were performed in a flow reactor at 493 K with a liquid feed rate of $0.2 \text{ cm}^3 \text{ min}^{-1}$, and a weight hourly space velocity (defined as grams of HMF per hour per gram of catalyst) of 0.15 h^{-1} and 0.98 h^{-1} for 1.5 wt% and 10 wt% HMF solutions in 1-butanol, respectively. CuRu/C catalysts were prepared by incipient wetness impregnation (defined as the condition where the impregnation liquid containing the metal salt is added to fill the pore volume of the catalyst support, equal to 1.06 ml per gram of catalyst in this case) of a commercial catalyst consisting of 10 wt% Ru on carbon (C-10 catalyst consisting of high performance ruthenium on Vulcan XC-72 carbon, E-TEK Division, PEMEAS Fuel Cell Technologies) with a copper nitrate ($\text{CuNO}_3 \cdot 2.5\text{H}_2\text{O}$, Sigma-Aldrich) water solution.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 December 2006; accepted 11 May 2007.

- Parikka, M. Global biomass fuel resources. *Biomass Bioenergy* **27**, 613–620 (2004).
- Ragauskas, A. J. et al. The path forward for biofuels and biomaterials. *Science* **311**, 484–489 (2006).
- Barlow, M. T., Smith, D. J. & Steward, D. G. Fuel composition. European patent EP0082689 (1983).
- Román-Leshkov, Y., Chheda, J. N. & Dumesic, J. A. Phase modifiers promote efficient production of hydroxymethylfurfural from fructose. *Science* **312**, 1933–1937 (2006).
- Brown, D. W., Floyd, A. J., Kinsman, R. G. & Roshan-Ali, Y. Dehydration reactions of fructose in non-aqueous media. *J. Chem. Technol. Biotechnol.* **32**, 920–924 (1982).
- Kuster, B. M. F. 5-Hydroxymethylfurfural (HMF). A review focussing on its manufacture. *Starch* **42**, 314–321 (1990).
- Moreau, C., Belgacem, M. N. & Gandini, A. Recent catalytic advances in the chemistry of substituted furans from carbohydrates and in the ensuing polymers. *Top. Catal.* **27**, 11–30 (2004).
- Szamt, H. H. & Chundury, D. D. The preparation of 5-hydroxymethylfurfuraldehyde from high fructose corn syrup and other carbohydrates. *J. Chem. Technol. Biotechnol.* **31**, 135–145 (1981).
- van Dam, H. E., Kieboom, A. P. G. & van Bekkum, H. The conversion of fructose and glucose in acidic media: Formation of hydroxymethylfurfural. *Starch* **38**, 95–101 (1986).
- Eisen, E. O. & Joffe, J. Salt effects in liquid-liquid equilibria. *J. Chem. Eng. Data* **11**, 480–484 (1966).
- Tan, T. C. & Aravindh, S. Liquid-liquid equilibria of water/acetic acid/1-butanol system—effects of sodium (potassium) chloride and correlations. *Fluid Phase Equil.* **163**, 243–257 (1999).
- Neier, W., Webers, W., Ruckhaber, R., Osterburg, G. & Ostwald, W. Continuous production of sec-butanol. German patent DE3040997 (1982).
- Jones, D. T. & Woods, D. R. Acetone-butanol fermentation revisited. *Microbiol. Rev.* **50**, 484–524 (1986).
- Ramey, E. D. Continuous two-stage dual path anaerobic fermentation of butanol and other organic solvents using two different strains of bacteria. US patent US5753474 (1998).
- Manly, D. G. & Dunlop, A. P. Catalytic hydrogenation. I. Kinetics and catalyst composition in the preparation of 2-methylfuran. *J. Org. Chem.* **23**, 1093–1095 (1958).
- Rao, R. S., Baker, R. T. & Vannice, M. A. Furfural hydrogenation over carbon-supported copper. *Catal. Lett.* **60**, 51–57 (1999).
- Zheng, H.-Y. et al. Towards understanding the reaction pathway in vapour phase hydrogenation of furfural to 2-methylfuran. *J. Mol. Catal. Chem.* **246**, 18–23 (2006).
- Twigg, M. V. & Spencer, M. S. Deactivation of supported copper metal catalysts for hydrogenation reactions. *Appl. Catal. A* **212**, 161–174 (2001).
- Helms, C. R. & Sinfelt, J. H. Electron spectroscopy (ESCA) studies of ruthenium-copper catalysts. *Surf. Sci.* **72**, 229–242 (1978).
- Sinfelt, J. H. Supported bimetallic cluster catalysts. *J. Catal.* **29**, 308–315 (1973).
- Sinfelt, J. H., Lam, Y. L. & Cusumano, J. A. Nature of ruthenium-copper catalysts. *J. Catal.* **42**, 227–237 (1976).
- Broughton, D. B. & Gerhold, C. G. Continuous sorption process employing fixed beds of sorbent and moving inlets and outlets. US Patent US2985589 (1961).
- Hashimoto, K., Adachi, S. & Shirai, Y. in *Preparative and Production Scale Chromatography* (eds Ganestos, G. & Barker, P. E.) 395–417 (CRC Publishing, New York, 1993).
- Jones, J. H., Fenske, M. R. & Rusk, R. A. Vapor-phase oxidation as a process for raising octane number. *Ind. Eng. Chem. Prod. Res. Dev.* **10**, 57–65 (1971).
- Graboski, M. S. *An Analysis of Alternatives for Unleaded Petrol Additives for South Africa* (Technical Report, United Nations Environment Programme, Nairobi, Kenya, 2003); available at <http://www.unep.org/PCFV/PDF/GraboskiReport.pdf>.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the National Science Foundation Chemical and Transport Systems Division of the Directorate for Engineering, and the US Department of Energy Office of Basic Energy Sciences. We thank R. McClain and the UW Chemistry Department for access to their mass spectrometer. We also thank D. Simonetti, R. West, J. Chheda, E. Kunke, S. Chen and S. Laumann for discussions and technical assistance.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A.D. (dumesic@engr.wisc.edu).

METHODS

Dehydration reactions. Aqueous- and organic-phase components including reactants (fructose and glucose), inorganic salts (NaCl, KCl, NaBr, KBr, NaNO₃, Na₂SO₄, Na₂HPO₄, CaCl₂, CsCl and MgCl₂), organic solvents (2-butanol, 1-butanol, MIBK, toluene and 1-hexanol), and the acid catalyst HCl were obtained from Sigma-Aldrich.

Batch catalytic experiments were carried out in 10 ml, thick-walled glass reactors (Alltech) heated in a temperature controlled oil bath placed on top of a magnetic stirrer. The temperature in the oil bath was measured by a K-type thermocouple (Omega) and controlled using a series 16A temperature controller (Dwyer Instruments) coupled with a 150 W heating cartridge (McMaster Carr). In a typical experiment, 1.5 g of 0.25 M HCl aqueous phase solution composed of 30 wt% fructose (salt-free basis) and an amount of extracting solvent necessary to keep $V_{\text{org}}/V_{\text{aq}} = 3.2$ were poured into the reactor. Before use, the extracting solvent was pre-contacted with an aqueous phase containing the same amount of salt as the aqueous phase in the dehydration reactor. In this way, the extracting phase was saturated with water and salt before reaction. The reactor was placed in a preheating oil bath set at 353 K for 2 min, and then transferred into an oil bath set at 453 K to perform the reaction. Reaction times ranged from 2.5 to 3 min to obtain conversions close to 75%. The reaction was stopped by cooling the reactor in an ethylene glycol bath at 253 K. Next, on separation, the masses and densities of both phases were measured. For each phase, density values were measured by weighing a constant volume.

Sample analyses were performed by HPLC using a Waters 2690 system equipped with PDA 960 UV (320 nm) and RI-410 refractive index detectors. Fructose disappearance was monitored with an Aminex HPX-87H column (Biorad), using MilliQ water (pH 2) as the mobile phase at a flow rate of 0.6 ml min⁻¹ and a column temperature of 333 K. HMF was quantified in the aqueous and organic phases with a Zorbax SB-C18 reverse phase column (Agilent), using a 2:8 v/v methanol:water (pH 2) gradient at a flow rate of 0.7 ml min⁻¹ and a column temperature of 308 K. The gradient is described in Supplementary Information.

Fructose conversion and HMF selectivity were calculated from the product of the aqueous and organic phase concentrations obtained in the HPLC and their corresponding volumes after reaction.

After the dehydration reaction, impurity levels of the organic layer were measured using a combination of techniques. Levels of NaCl were measured using the Aminex HPX-87H column in the HPLC; acid content was measured by acid titration with NaOH; and 1-chlorobutane content was measured with gas chromatography (Shimadzu GC-2010 with an FID detector and a DB-5 column from Alltech). Water content before and after the purification step was obtained by mass balance. Specifically, given that the initial amounts of all components are known, as well as the final masses of the aqueous and organic phases and the concentration of the organic solvent in the aqueous layer (HPLC), it is possible to calculate the water content in the extracting solvent. For example, after mixing 5 g of water saturated with NaCl with 10 g of anhydrous 1-butanol, the mass of the organic layer is 10.6 g and it is determined that 0.06 g of 1-butanol is present in the aqueous layer. Then, assuming that a negligible amount of NaCl transfers into the organic layer, the water content in the organic layer is 0.66 g of H₂O in 10.6 g of extracting solvent or ~2,800 mmol of H₂O per l of extracting solvent. In a similar fashion, the amount of water remaining in the extracting solvent after the purification step can be calculated by subtracting the amount of water recovered from the evaporator from the initial amount of water present in the extracting solvent before evaporation.

Hydrogenolysis reactions: catalyst preparation. CuRu/C catalysts were prepared by incipient wetness impregnation of a commercial catalyst consisting of 10 wt% Ru on carbon (C-10: HP ruthenium on Vulcan XC-72, E-TEK Division, PEMEAS Fuel Cell Technologies) with a copper nitrate (CuNO₃·2.5H₂O, Sigma-Aldrich) water solution. For a typical batch of 3:2 (molar ratio) Cu:Ru catalyst, 1.55 g of copper nitrate was dissolved in 5 g of deionized (DI) water. This solution was then added drop-wise to 4.58 g of Ru/C catalyst. Following impregnation, the catalyst was dried in air at 403 K for 2 h and reduced at 523 K in flowing hydrogen for 10 h (0.42 K min⁻¹ ramp for 6 h followed by 4 h at 523 K). After reduction, the catalyst was allowed to cool to room temperature and passivated in flowing 2% oxygen in helium for 3 h. All gas flow rates were maintained at approximately 110 cm³(STP) min⁻¹. Pre-reduced barium promoted CuCrO₄ was used untreated from Sigma-Aldrich.

Hydrogenolysis reactions: batch reactor system. All batch reactor runs were carried out using an autoclave reactor with external temperature and stirring controller (Models 4566 and 4836, Parr Instrument). For a typical hydrogenolysis run, 2.5 g of HMF (98%, Sigma-Aldrich) was dissolved in 47.5 g of organic solvent. The solvent used was either dry 1-butanol (99.9%, Sigma-Aldrich) or 1-butanol pre-contacted with a NaCl/water solution that simulated the final untreated organic layer from the biphasic fructose dehydration step. The

NaCl/water solution was made by adding 6.7 g sodium chloride into 18.9 g DI water. Next, 51 g of 1-butanol was added to the NaCl/water solution and shaken vigorously. The resulting two phases were allowed to separate for 20 min. Afterwards the organic layer was siphoned off and used as the solvent. Next, 0.75 g of CuRu/C catalyst was added to the reactor. The reactor was sealed and purged of air by adding and releasing hydrogen to a pressure of 20 bar. Hydrogenolysis reactions were carried out at 493 K with 6.8 bar initial hydrogen pressure for 10 h while using a stirring speed of 400 r.p.m. These conditions were found to be optimal for DMF yield. After 10 h the reactor was cooled to room temperature before its contents were sampled, filtered (using 0.2 µm PES syringe membrane filter), and analysed.

Hydrogenolysis reactions: flow reactor system. Supplementary Fig. 2 shows the down-flow, vapour-phase, fixed-bed reactor set-up used to convert HMF to DMF. One gram of catalyst in powder form was mixed with 2.3 g of silicon dioxide fused granules with a 4–16 mesh size (Aldrich) and loaded into a 1/4-inch outer diameter tubular stainless steel reactor. The catalyst bed was contained in the tubular reactor by an end-plug of quartz wool (Alltech). A type-K thermocouple (Omega) attached to the outside of the reactor was used to measure the reactor temperature, which was controlled with a 16A series temperature controller (Dwyer Instruments). The flow rate of H₂ was controlled with a mass-flow meter (5850 Brooks Instruments). An HPLC pump (Model 301, Alltech) was used to introduce the feed solution into the down-flow reactor through a needle. The effluent from the reactor was condensed at room temperature in a separator, allowing for periodic sampling of the liquid product stream. The effluent gas stream passed through a back-pressure regulator (GO Regulator, Model BP-60), which controlled the system pressure, and through a flow meter to measure the gas flow rate.

All runs were carried out at 100% conversion at a temperature of 493 K, using a liquid feed rate of 0.2 cm³ min⁻¹, and a weight hourly space velocity (defined as grams of HMF per hour per gram of catalyst) of 0.15 h⁻¹ and of 0.98 h⁻¹ for 1.5 and 10 wt% runs. Other process conditions used in the experiments are listed in Supplementary Table 2. Product sampling took place for approximately every 3–6 cm³ of liquid feed, and reported values are mean values over all steady state points.

LETTERS

The early Miocene onset of a ventilated circulation regime in the Arctic Ocean

Martin Jakobsson¹, Jan Backman¹, Bert Rudels³, Jonas Nycander², Martin Frank⁴, Larry Mayer⁵, Wilfried Jokat⁶, Francesca Sangiorgi⁷, Matthew O'Regan⁸, Henk Brinkhuis⁷, John King⁸ & Kathryn Moran⁸

Deep-water formation in the northern North Atlantic Ocean and the Arctic Ocean is a key driver of the global thermohaline circulation and hence also of global climate¹. Deciphering the history of the circulation regime in the Arctic Ocean has long been prevented by the lack of data from cores of Cenozoic sediments from the Arctic's deep-sea floor. Similarly, the timing of the opening of a connection between the northern North Atlantic and the Arctic Ocean, permitting deep-water exchange, has been poorly constrained. This situation changed when the first drill cores were recovered from the central Arctic Ocean². Here we use these cores to show that the transition from poorly oxygenated to fully oxygenated ('ventilated') conditions in the Arctic Ocean occurred during the later part of early Miocene times. We attribute this pronounced change in ventilation regime to the opening of the Fram Strait. A palaeo-geographic and palaeo-bathymetric reconstruction of the Arctic Ocean, together with a physical oceanographic analysis of the evolving strait and sill conditions in the Fram Strait, suggests that the Arctic Ocean went from an oxygen-poor 'lake stage', to a transitional 'estuarine sea' phase with variable ventilation, and finally to the fully ventilated 'ocean' phase 17.5 Myr ago. The timing of this palaeo-oceanographic change coincides with the onset of the middle Miocene climatic optimum³, although it remains unclear if there is a causal relationship between these two events.

The Integrated Ocean Drilling Program (IODP) Expedition 302 (ACEX) in 2004 cored a 428-m-thick sediment sequence from the crest of the Lomonosov ridge in the central Arctic Ocean² (Fig. 1). The sediments recovered from two neighbouring drill sites permit the first geological 'ground truth' validation of the Cenozoic palaeoenvironmental history of the central Arctic Ocean⁴, which previously was based solely on interpretation of geophysical data^{4–6}. Evidence for the onset of a ventilated circulation system in the central Arctic Ocean is preserved by distinct, documentable changes in the chemical and physical properties and micropalaeontology of the recovered seafloor sediments. Neogene and Quaternary sediments from the Lomonosov ridge are found within a single lithostratigraphic unit, subdivided into six subunits².

This study focuses on the interval around subunit 1/5, a 5.76-m-thick section between 193 and 199 metres composite depth (m.c.d.) containing a sequence of alternating grey and black layers (Fig. 2). The uppermost grey layers of this subunit are marked by a sharp colour change, with dark brown sediments above⁷. The black layers of subunit 1/5 contain much higher total organic carbon (TOC) contents (4.7–14.1%) than the lighter intervening grey layers (0.1–3.0%)⁸. Below this unit, TOC values are consistently 1–5%, and above they are <0.5% (ref. 2). Samples from the greyish layers in

subunit 1/5 have an oxic character, whereas samples from dark layers have a euxinic character⁸. These changes in sediment properties all indicate that the interval around subunit 1/5 represents a time of transition from a poorly ventilated and land-locked sea with reduced salinities to a generally well-ventilated saline ocean. This transition is further supported by the sudden appearance of benthic agglutinated foraminifera in the brown interval directly above subunit 1/5 (lower subunit 1/4, 187.4 m.c.d.)¹. Finally, the general lack of palynomorphs in subunit 1/4 and low TOC contents is consistent with oxic depositional conditions⁹. Taken together, these changes document that subunit 1/5 preserved the transition from euxinic to oxic conditions in the central Arctic Ocean (Fig. 2).

The timing of this transition is established by taking advantage of the fact that subunit 1/5 is unique in that it contains a monotypic assemblage of an abundant peridinioid taxon of Burdigalian age¹⁰. In the absence of more precise age information, we have used the mid-point of the late early Miocene (Burdigalian), 18.2 Myr ago¹¹, for the base of subunit 1/5. Sedimentation rates estimates suggest that this 5.76-m-thick subunit represents about 0.75 ± 0.1 million years of deposition, and that its top has an age of about 17.5 Myr when 18.2 Myr is used as the age for the base of the unit (see Methods for age model and Supplementary Information for uncertainties).

In order to place our core results in a wider regional context, the sediment sequence that preserves this critical palaeo-oceanographic change was correlated with a seismic profile (AWI-91090), which was used to locate the IODP drill site (Fig. 1). The changes in physical properties of the sediments between subunits (1/6 to 1/4) are large enough that they can, through modelling, be related to seismic reflections (Fig. 2) (see Methods for seismic modelling). These reflectors can, in turn, be correlated with a seismostratigraphic subdivision (LR-3 to LR-6, where LR-6 is the uppermost unit) already established for the Lomonosov ridge sediment sequence¹². The boundary between seismostratigraphic units LR-5 and LR-4 represents a set of reflectors that correlates both to the transition from ACEX unit 2 to subunit 1/6 (density controlled) and from subunit 1/6 to subunit 1/5 (velocity controlled) just preceding the early Miocene initiation of ventilated circulation in the Arctic Ocean (Fig. 2). The two subunits are separated by a hiatus. The set of reflectors defining the LR-5 and LR-4 boundary can be traced regionally on the Lomonosov ridge, and can also be linked to the seismic stratigraphy of the Amundsen and Makarov basins as well as the East Siberian continental margin slope^{12,13}. Thus, the core obtained from the Lomonosov ridge reflects circulation changes that occurred over large parts of the Arctic Ocean.

To better understand the context of this crucial oceanographic transition, we have reconstructed the palaeogeography and

¹Department of Geology and Geochemistry, ²Department of Meteorology, Stockholm University, SE-106 91 Stockholm, Sweden. ³Finnish Institute for Marine Research, F-00561 Helsinki, Finland. ⁴Leibniz Institute of Marine Sciences at the University of Kiel (IFM-GEOMAR), DE-24148 Kiel, Germany. ⁵Center for Coastal and Ocean Mapping, University of New Hampshire, Durham, New Hampshire 03824, USA. ⁶Alfred Wegener Institute for Polar Research, DE-27570 Bremerhaven, Germany. ⁷Laboratory of Palaeobotany and Palynology, Utrecht University, NL-3584 Utrecht, The Netherlands. ⁸Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882, USA.

palaeo-bathymetry of the region during the early Miocene (see Methods and Supplementary Information). Major Cenozoic deltaic successions mapped in the Beaufort Mackenzie basin indicate continuous input of large volumes of fresh water into a land-locked Arctic Ocean basin¹⁴. Our palaeo-geographic and palaeo-bathymetric map for the late early Miocene suggests that the only significant outlet for Arctic Ocean fresh water was to the North Atlantic through the Fram Strait (Fig. 3). Straits are physiographic bottlenecks, strongly affecting the circulation by confining broad current flows^{15,16}. The exchange through the present, >400 km wide, Fram Strait is sufficient to allow all layers in the Arctic Ocean to be ventilated through advection, down to the sill depth (2,550 m; ref. 17). Ventilation below that depth is at present driven by shelf-slope convection¹⁸. During Miocene times, when the Fram Strait opened and deepened through sea-floor spreading, the water exchange between the Arctic and North Atlantic must have developed through a series of changes that also influenced the upstream basin circulation and ventilation conditions within the Arctic Ocean.

The initial opening phase of the Fram Strait was restricted to a uni-directional hydraulically controlled freshwater outflow (Fig. 4a). During this ‘Arctic lake’ stage, ventilation of deep waters may have occurred via seasonal convection. We hypothesize that when the

Fram Strait widened and deepened, a compensating inflow of saline North Atlantic water became possible, resulting in a bi-directional, two-layer flow through the strait (Fig. 4b). The Arctic then evolved from its lake-stage to an enclosed estuarine sea, much like the modern Black Sea. Anoxic conditions prevail in the Black Sea because of strong stable stratification and limited exchange through the 36–124 m deep and <4 km wide Bosphorus Strait, where inflowing saline Mediterranean waters comprise the main source of ventilated waters to the deep Black Sea¹⁹. The volume of the Arctic Ocean and its present net freshwater discharge both are about 20 times greater compared to the Black Sea. The present in- and outflow water through the Fram Strait is about 10 Sv, with a net outflow of 2 Sv. This is 1,000 times larger than the exchanges through the Bosphorus, and the low-salinity surface outflow is about 1 Sv, 50 to 100 times that of the Bosphorus.

A possible explanation for the alternations between euxinic conditions (dark layers) and intervening more oxic conditions (grey layers), in the transitional subunit 1/5, is oscillations in sea level. Changes in sea level will create an ‘on–off switch’ for Arctic Ocean circulation: during low sea level, there would be ‘lake stages’ having seasonal convection and uni-directional one-layer outflow; during times of high sea level, there would be estuarine, ‘Black

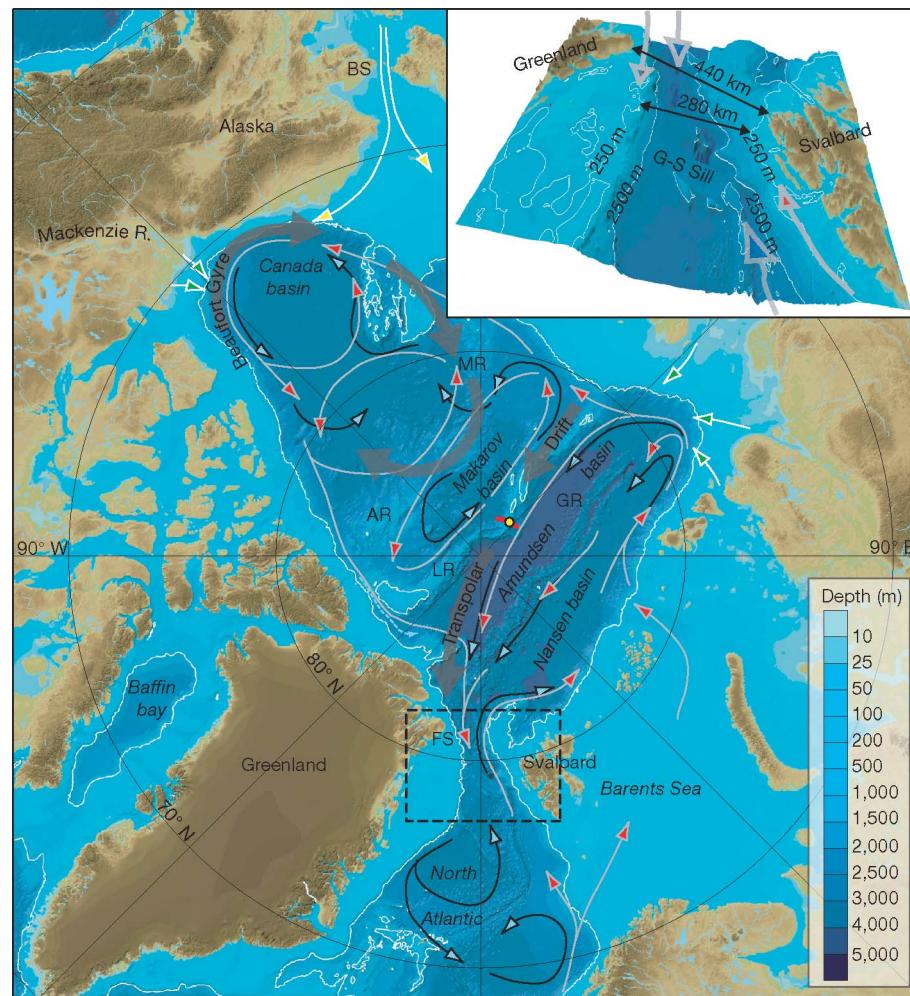


Figure 1 | Schematic map of the present ocean circulation in the Arctic Ocean. Shown are the present circulation of intermediate waters of Atlantic origin (grey arrows, red arrowheads), and deep waters (black arrows, light blue arrowheads), updated from ref. 18, and major freshwater inputs by rivers (white arrows, green arrowheads); also shown is the Pacific water influx through the Bering Strait (white arrows, yellow arrowheads). ACEX coring site, yellow circle; seismic reflection profile AWI-91090¹², red line. Bathymetry is from the International Bathymetric Chart of the Arctic Ocean

(IBCAO)²⁰. The white contour line represents the 1,000 m isobath. Physiographic features: AR, Alpha ridge; BS, Bering Strait; FS, Fram Strait; GR, Gakkel ridge; LR, Lomonosov ridge; MR, Mendelev ridge. Inset, present seafloor morphology of the Fram Strait (dashed box in main figure) based on IBCAO updated with multibeam bathymetry¹⁷. Shortest distances between Svalbard’s and Greenland’s 250 m isobaths, and coastlines, are shown. G-S, Greenland–Spitsbergen sill. Arrows indicate generalized water mass exchange between the Arctic Ocean and the North Atlantic.

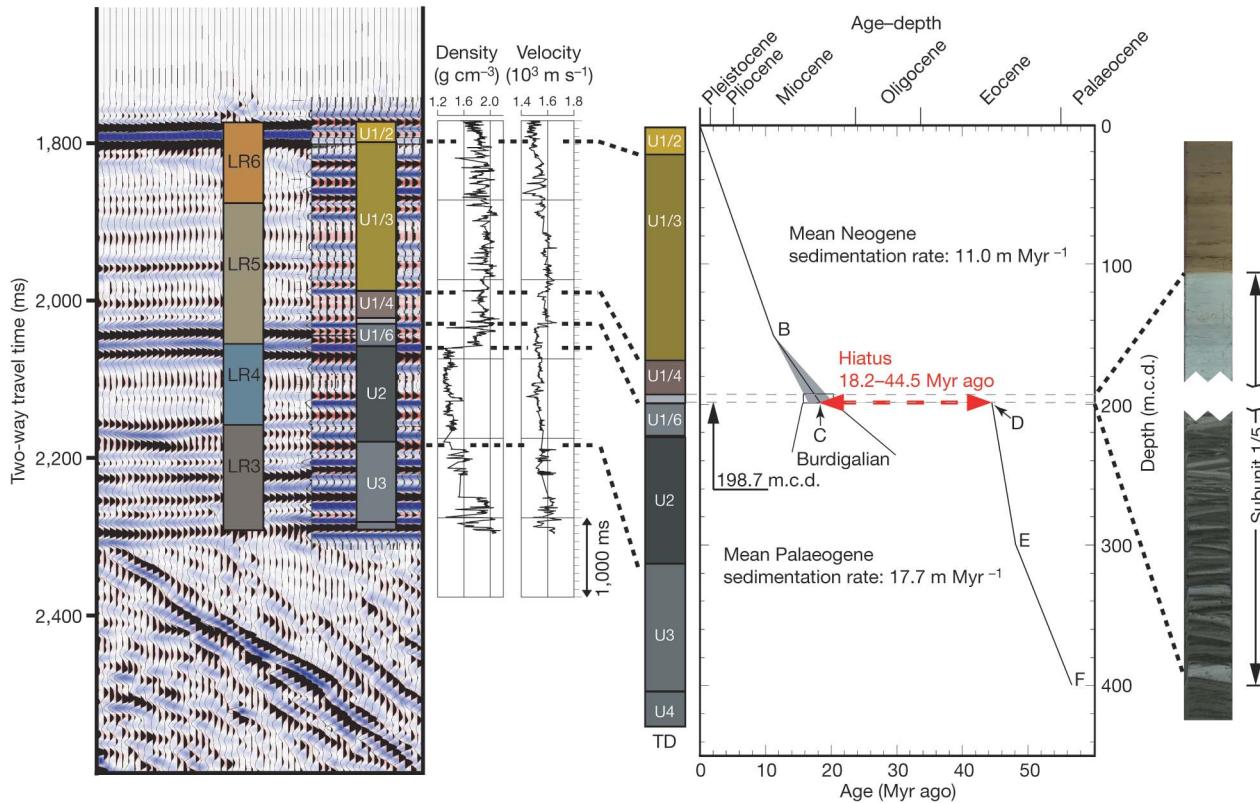


Figure 2 | Correlation between a synthetic seismogram representing the stratigraphy of the ACEX sites and seismic reflection profile AWI-91090, and the ACEX age model. The shown section of seismic reflection profile AWI-91090 crosses Site M0004². Stratigraphic units U1–U4 are as inferred by ref. 2, and seismic units LR6–LR3 are as inferred in ref. 12. See Methods

for age model and Supplementary Information for age uncertainties. Measured density and velocity records are shown next to the synthetic seismogram, which is re-sampled in two-way travel time. Age control points (Myr ago): B, 12.3; C, 18.2; D, 44.4; E, 48.6; and F, 55 (see Supplementary Information). TD, terminal depth.

Sea like' conditions, with two-layer, bi-directional flow. Sea-level varied between about 15 and 30 m during late early Miocene times²⁰.

As the strait deepened further, the sea-level changes were no longer sufficient for a reversal to 'Arctic lake' conditions. Estuarine conditions would then prevail, with hydraulic control of the outflow of low-salinity water in the upper layer, and deep inflow of saline Atlantic water. The final transition occurred when the strait became wider than the internal Rossby radius of the upper layer. In the final regime, the outflow in the upper layer was rotationally controlled, rather than hydraulically controlled²¹.

We have considered the scaling of the estuarine circulation using a simple analytic two-layer model (Supplementary Information). In this model, the circulation is forced by the wind-driven vertical mixing and by the freshwater influx from rivers to the Arctic basin. We find that the strait width at which the transition from hydraulic control to rotational control occurs is independent of the mixing, which is the most uncertain parameter, and that for the present freshwater forcing of about 0.2 Sv the transition width is 13 km. (Note, however, that the transition between the two regimes is gradual, rather than sharp.)

In the rotationally controlled regime, the outflow of low-salinity surface water is concentrated at the western (Greenland) continental slope of the strait, as is the case today (Fig. 4c). This opens the eastern part of the strait for flow driven by mechanisms other than turbulent entrainment in the Arctic Ocean interior—that is, barotropic currents, attached to the bathymetry and driven by the large-scale wind field, can enter the Arctic Ocean. The total present inflow via the Fram Strait is 5 to 10 times greater than the outflow of low-salinity surface water²². Much of this exchange represents recirculation within the strait, although 4–5 Sv, at least, enters the Arctic Ocean and will contribute to ventilating its deep waters^{22,23}.

At what point in this development did the deep circulation become strong enough to prevent anoxic conditions in the Arctic

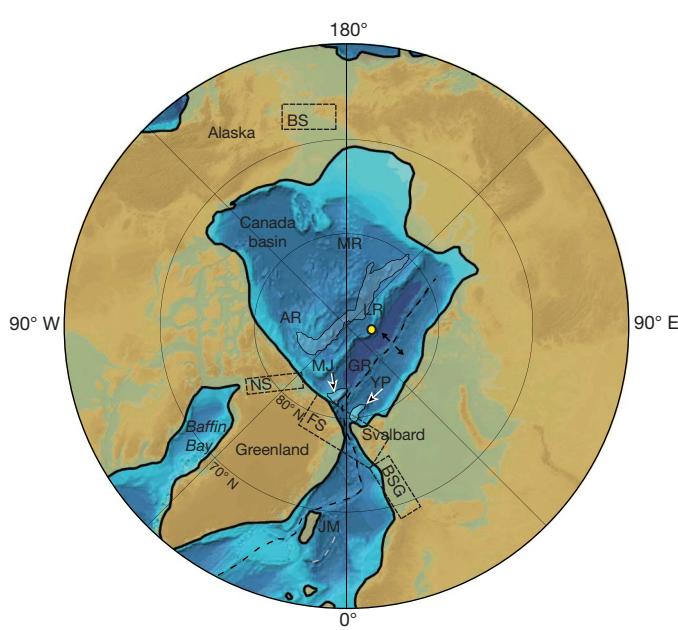


Figure 3 | Palaeo-geographic/palaeo-bathymetric reconstruction for the late early Miocene. Symbols and abbreviations as Fig. 1, with the addition of: BSG, Barents Sea gateway; JM, Jan Mayen microcontinent; KR, Knipovich ridge; MJ, Morris Jessup rise; NS, Nares Strait; YP, Yermak plateau. See Methods for compilation approach.

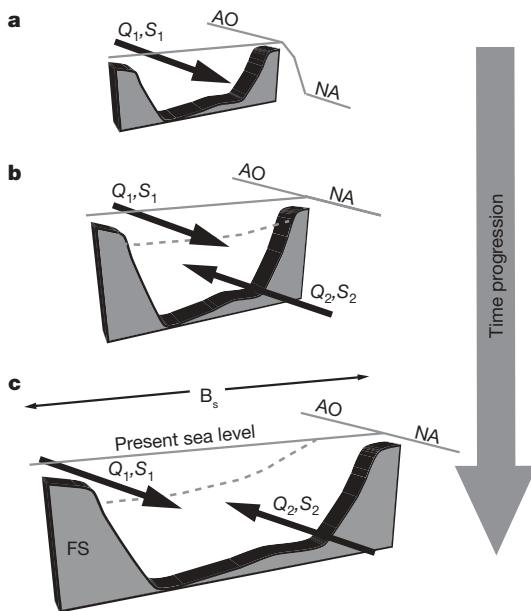


Figure 4 | A schematic illustration of the Fram Strait opening and hypothetical water exchange development between the Arctic Ocean and North Atlantic. AO, Arctic Ocean; NA, North Atlantic. **a**, A narrow strait resulting in a uni-directional hydraulically controlled outflow from the Arctic. S_1 is the salinity of the out-flowing flux of water Q_1 . **b**, A wider and deeper strait allowing the establishment of a bi-directional, two-layer flow through the strait due to a compensating inflow (Q_2) of saline (S_2) North Atlantic water. This phase in the Arctic's palaeo-oceanographic development is analogous to the present Black Sea. **c**, The Fram Strait becomes wide enough that the influence of the Earth's rotation changes the water flow through the strait to a rotationally controlled bi-directional two-layer flow. This opens the possibility of a barotropic current flow through the strait. Our scaling analysis (Supplementary Information) suggests that this turning point is independent of the mixing and can, assuming the present freshwater input of 0.2 Sv, occur at a strait width of 13 km. At present the strait width (B_s) is >400 km for the Fram Strait.

basin? Our scaling analysis indicates that with today's wind speed and freshwater forcing, this may have occurred before the transition to the rotationally controlled regime. However, this result is very sensitive to the wind speed, which is unknown during the early Miocene. Moreover, the analysis neglects processes (for example, shelf convection) that would tend to strengthen the deep stratification, and thereby inhibit the ventilation of the deep waters. It is therefore likely that oxygenated conditions were not established until a rotationally controlled regime had been established, when the strait was wide enough for barotropic currents to meet there. For this, a strait width of 40–50 km should be sufficient (Supplementary Information).

According to the age model used, the change in ventilation history for the Arctic Ocean began during the late early Miocene about 18.2 Myr ago and was completed about 0.7 Myr later, about 17.5 Myr ago. The Arctic Ocean seafloor spreading had propagated at least as far south as 81° N at 16.2 Myr ago (chron C5Cn.1), according to new seismic reflection and aeromagnetic data²⁴. According to the age-width estimation using these geophysical data (Supplementary Fig. S5), the Fram Strait began to open at great depths (present-day water depth >2,000 m) by 13.7 Myr ago, that is, 3.8 Myr after the Arctic Ocean had gone through its transitional phase into a ventilated circulation, as marked by the end of subunit 1/5. However, an initial corridor of immature seafloor spreading may have developed between Greenland and Svalbard before the full spreading extended through this gateway²⁵, and it is difficult to quantify the volumetric exchange of water with the North Atlantic on the basis of this estimate.

METHODS SUMMARY

The ACEX age model is based on ^{10}Be stratigraphy in the upper 151 m.c.d. (ref. 7) and biostratigraphy from 151 m.c.d. to 400 m.c.d. (ref. 2). A single geomagnetic reversal boundary is used (top chron C25n). The timescale used is compiled as follows: Neogene¹¹; top chron C6n.3n to base chron C19n²⁶; top chron 20n and older²⁷. Our palaeo-geographic and palaeo-bathymetric reconstruction is based on plate tectonic maps generated from the Ocean Drilling Stratigraphic Network (ODSN) tools available online at <http://www.odsn.de>. Lithosphere plates were moved to 18 Myr ago (mid-point of the Burdigalian) relative to the hotspot reference frame of ref. 28. To derive the palaeo-bathymetry, the International Bathymetric Chart of the Arctic Ocean (IBCAO)²⁹ Digital Terrain Model (DTM) was rectified to fit the plate tectonic reconstruction using the coast line as reference. Published information was used to infer details about eustatic sea level, location of palaeo-shorelines, uplift or subsidence of regions, and opening of gateways (Supplementary Fig S2). Sea level data were adopted from ref. 20. Synthetic seismic modelling assuming planar waves, with no multiples or signal attenuation, has been performed using software by Divest. To simulate the seismic reflection profile AWI-91090 crossing the ACEX sites, a Ricker wavelet with a peak frequency of 40 Hz and a period of 19.5 ms was used as source function for convolution with the ACEX core impulse response (reflectivity function), which was calculated using the bulk density and P-wave velocity records. The reflectivity function was sampled at 0.25 ms before convolution with the Ricker wavelet and the generation of synthetic traces. Logged P-wave velocities were not corrected to simulate *in situ* conditions with respect to water depth (pressure) and temperature. Nor were any corrections for porosity rebound applied owing to the dominantly terrigenous composition of the sediment. Details can be found in Supplementary Information.

Received 22 March; accepted 10 May 2007.

1. Agaard, K. & Carmack, E. C. in *The Polar Oceans and their Role in Shaping the Global Environment: The Nansen Centennial Volume* (eds Johannessen, O. M., Muench, R. D. & Overland, J. E.) 5–20 (Geophysical Monograph 85, American Geophysical Union, Washington DC, 1994).
2. Backman, J., Moran, K., McInroy, D. B., Mayer, L. A. & the Expedition 302 Scientists. Expedition 302 Summary. *Proc. IODP* 302 doi:10.2204/iodp.proc.302.101.2006 (Integrated Ocean Drilling Program Management International, College Station, Texas, 2006).
3. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292, 686–693 (2001).
4. Lawver, L. A. & Scotese, C. R. in *The Arctic Ocean Region, Geology of North America* (eds Grantz, A., Johnson, G. L. & Sweeney, J. F.) Vol. L, 593–618 (GSA, Boulder, Colorado, 1990).
5. Brozena, J. M. et al. New aerogeophysical study of the Eurasia Basin and Lomonosov Ridge: Implications for basin development. *Geology* 31, 825–828 (2003).
6. Vogt, P. R., Taylor, P. T., Kovacs, L. C. & Johnson, G. L. Detailed aeromagnetic investigation of the Arctic Basins. *J. Geophys. Res. B* 84, 1071–1089 (1979).
7. Moran, K. et al. The Cenozoic palaeoenvironment of the Arctic Ocean. *Nature* 441, 601–605 (2006).
8. Stein, R., Boucsein, B. & Meyer, H. Anoxia and high primary production in the Paleogene central Arctic Ocean: First detailed records from Lomonosov Ridge. *Geophys. Res. Lett.* 33, L18606, doi:10.1029/2006GL026776 (2006).
9. Batten, D. J. in *Palyontology: Principles and Applications* (eds Jansoni, J. & McGregor, D. C.) Vol. 3, 1021–1064 (American Association of Stratigraphic Palyngologists Foundation, Salt Lake City, 1996).
10. Williams, G. L. & Manum, S. B. Oligocene-early Miocene dinocyst stratigraphy of Hole 985A (Norwegian Sea). *Proc. ODP Sci. Res.* 162, 99–109 (1999).
11. Lourens, L., Hilgen, F., Shackleton, N. J., Laskar, J. & Wilson, D. in *A Geologic Time Scale 2004* (eds Gradstein, F., Ogg, J. & Smith, A.) 409–440 (Cambridge Univ. Press, Cambridge, UK, 2004).
12. Jokat, W., Weigelt, E., Kristoffersen, Y., Rasmussen, T. & Schöne, T. New insights into the evolution of the Lomonosov Ridge and the Eurasian Basin. *Geophys. J. Int.* 122, 378–392 (1995).
13. Langinen, A. E., Gee, D. G., Lebedeva-Ivanova, N. N. & Zamansky, Y. Y. Velocity structure and correlation of the sedimentary cover on the Lomonosov Ridge and in the Amerasian Basin, Arctic Ocean. (Fourth Int. Conf. on Arctic Margins, ICAM IV, Dartmouth, Nova Scotia, Canada, 2006); (<http://www.mms.gov/alaska/icam/>).
14. Dixon, J., Dietrich, J. R. & McNeil, D. H. Upper Cretaceous to Pleistocene sequence stratigraphy of the Beaufort-Mackenzie delta and banks areas, northwest Canada. *Geol. Surv. Can. Bull.* 407, 1–52 (1992).
15. Lane-Serff, G. F. Topographic and boundary effects on steady and unsteady flow through straits. *Deep-sea Res. II* 51, 321–334 (2004).
16. Pratt, L. J. Recent progress on understanding the effects of rotation in models of sea straits. *Deep-sea Res. II* 51, 351–369 (2004).
17. Klenke, M. & Schenke, H. W. A new bathymetric model for the central Fram Strait. *Mar. Geophys. Res.* 23, 367–378 (2002).
18. Rudels, B., Jones, E. P., Anderson, L. G. & Kattnar, G. in *The Polar Oceans and their Role in Shaping the Global Environment* (eds Johannessen, O. M., Muench, R. D. &

Overland, J. E.) 33–46 (Geophysical Monograph Vol. 85, American Geophysical Union, Washington DC, 1994).

19. Stanev, E. V. Understanding Black Sea dynamics. *Oceanography* **18**, 56–75 (2005).
20. Miller, K. G. et al. The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298 (2005).
21. Whitehead, J. A., Leetmaa, A. & Knox, R. A. Rotating hydraulics of strait and sill flows. *Geophys. Fluid Dyn.* **6**, 101–125 (1974).
22. Fahrbach, E. et al. Direct measurements of volume transports through Fram Strait. *Polar Res.* **20**, 217–224 (2001).
23. Schauer, U., Fahrbach, E., Österhus, S. & Rohard, G. Arctic warming through the Fram Strait — oceanic heat transport from three years of measurements. *J. Geophys. Res. C* **109**, C06026, doi:10.1029/2003JC001823 (2004).
24. Jokat, W., Leinweber, V., Ehlers, B. M., Boebel, T. & Schenke, H. W. Timing and geometry of the Fram Strait opening. *Geophys. J. Int.* (submitted).
25. Engen, Ø. *Evolution of High Arctic Ocean Basins and Continental Margins*. Thesis, Univ. Oslo (2005).
26. Pälike, H. et al. The heartbeat of the Oligocene climate system. *Science* **314**, 1894–1898 (2006).
27. Cande, S. C. & Kent, D. V. Revised calibration of the geomagnetic polarity timescale for the Late Cretaceous and Cenozoic. *J. Geophys. Res. B* **100**, 6093–6095 (1995).
28. Müller, R. D., Royer, J.-Y. & Lawner, L. A. Revised plate motions relative to the hotspots from combined Atlantic and Indian Ocean hotspot tracks. *Geology* **21**, 275–278 (1993).
29. Jakobsson, M., Cherkis, N., Woodward, J., Macnab, R. & Coakley, B. New grid of Arctic bathymetry aids scientists and mapmakers. *Eos* **81**, 89, –93, 96 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature/.

Acknowledgements Financial support was received from Stockholm University, the Swedish Research Council (VR), the Swedish Royal Academy of Sciences through a grant financed by the Knut and Alice Wallenberg Foundation, and the Netherlands Organization for Scientific Research (NWO). The ACEX expedition was carried out by the Integrated Ocean Drilling Program (IODP). We thank the IODP European Science Operator (ESO), the Swedish Polar Research Secretariat as well the ship fleet management, captains and crews of the icebreakers *Oden*, *Vidar Viking* and *Sovetskiy Soysuz*. N. Welters is thanked for sample processing. Comments and suggestions from M. Siddall are acknowledged. This is a contribution from the Stockholm University Climate Research Centre (SUCLIM).

Author Contributions M.J. and J.B. initiated the paper. M.J. compiled the palaeo-bathymetric reconstruction, performed the core-seismic integration and took part in the development of the oceanographic analysis, which was led by B.R. and J.N. The age model was developed by J.B., M.F provided ^{10}Be data, H.B and F.S. provided micropalaeontological information, and J.K provided palaeointensity data. W.J. compiled the age-width estimation for the Fram Strait from geophysical data. All authors discussed the results and provided input to the manuscript during its development.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.J. (martin.jakobsson@geo.su.se).

LETTERS

Electromagnetic detection of a 410-km-deep melt layer in the southwestern United States

Daniel A. Toffelmier¹ & James A. Tyburczy¹

A deep-seated melt or fluid layer on top of the 410-km-deep seismic discontinuity in Earth's upper mantle, as proposed in the transition-zone 'water filter' hypothesis¹, may have significant bearing on mantle dynamics and chemical differentiation. The geophysical detection of such a layer has, however, proved difficult. Magneto-telluric and geomagnetic depth sounding are geophysical methods sensitive to mantle melt. Here we use these methods to search for a distinct structure near 410-km depth. We calculate one-dimensional forward models of the response of electrical conductivity depth profiles, based on mineral physics studies of the effect of incorporating hydrogen in upper-mantle and transition-zone minerals. These models indicate that a melt layer at 410-km depth is consistent with regional magnetotelluric and geomagnetic depth sounding data from the southwestern United States (Tucson)². The 410-km-deep melt layer in this model has a conductance of 3.0×10^4 S and an estimated thickness of 5–30 km. This is the only regional data set that we have examined for which such a melt layer structure was found, consistent with regional seismic studies³. We infer that the hypothesized transition-zone water filter¹ occurs regionally, but that such a layer is unlikely to be a global feature.

The presence of a melt layer at the base of the upper mantle (410-km depth), as proposed in the transition-zone water filter hypothesis¹, would control water and trace element abundances in the upper mantle as well as indicate a hydrated transition zone. This hypothesis may also explain the differing chemical compositions of ocean island basalts and mid-ocean-ridge basalts¹. Teleseismically imaged low (seismic) velocity zones have been reported in several areas, and may indicate the existence of a thin melt or fluid layer at this depth in the upper mantle^{3,4}. Song *et al.*³ observe a shear velocity decrease of as much as 5% between 320-km and 410-km depth beneath the western United States. Detection and characterization of this layer could have significant bearing on mantle dynamics, formation and hydration.

Magnetotelluric/geomagnetic depth sounding (MT/GDS) interpretation using mineral physics based electrical conductivity (σ) depth profiles offers the potential for resolving deep upper-mantle features. Using recent measurements on effects of H₂O on the electrical conductivity of upper-mantle and transition-zone minerals^{5,6} and melt considerations^{7–9}, we construct mineral-physics-based electrical conductivity depth profiles of the upper mantle and transition zone consistent with seismic constraints, including the effects of water and a 410-km-deep melt/fluid layer. Forward modelling of the MT/GDS response of these one-dimensional (1D) electrical conductivity depth profiles facilitates characterization of the Earth's electrical response to a 410-km-deep melt layer, and allows comparison to regional MT/GDS field data. Five sufficiently long period ($\geq 10^7$ s), 1D regional MT/GDS data sets were considered in this study (the southern Basin and Range², the French Alps¹⁰, North Pacific Ocean¹¹, a European average¹², and the Canadian Shield¹³).

Electrical conductivity depth profiles are forward modelled following the methods of ref. 14, and are compared to regional MT/GDS data as apparent resistivity, $\rho(\tau)$, and impedance phase, $\varphi(\tau)$, as functions of period, τ . $\rho(\tau)$ and $\varphi(\tau)$ are computed from measured electric and magnetic fields (MT/GDS response) through the complex impedance tensor¹⁵. The initial electrical conductivity depth profile is calculated using the mineralogy and dry conductivity values of ref. 14 and the geotherm of ref. 16 (see Methods). Then we vary the upper-mantle water content in olivine⁵ and the transition zone⁶ to minimize the misfit between regional data and forward models, yielding the melt-free regional model. If warranted by transition-zone water content, a high-conductivity (melt) layer is introduced near 410-km depth and the resulting improvement or degradation of the fit is examined.

To determine the sensitivity of the method to a deep upper-mantle melt zone, the MT/GDS response of a 1-km-thick high-conductivity layer of varying conductivity inserted into the initial model is analysed (Fig. 1). The effects of the high-conductivity layer are seen most

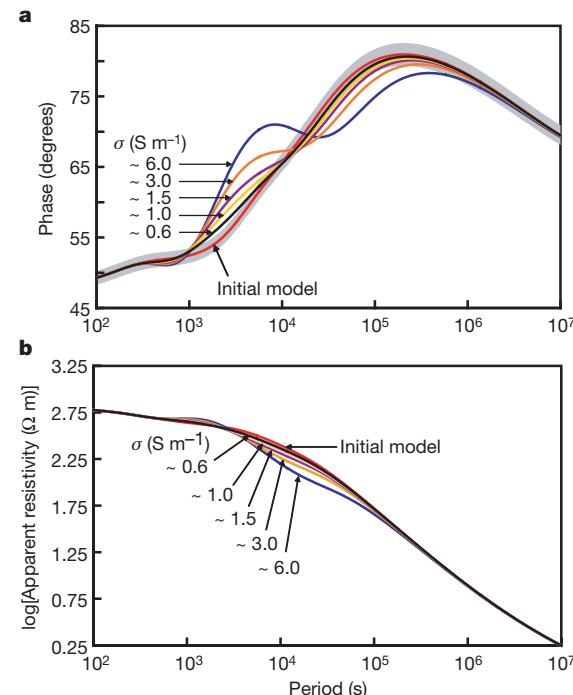


Figure 1 | Forward modelled MT response of upper-mantle conductivity depth profile. Model includes 1-km-thick melt layer at 410-km depth with varying melt electrical conductivity (σ), and illustrates the deviations from the melt-free (initial) model. **a**, $\varphi(\tau)$ models with $\pm 2\%$ error (grey region) on the initial model, and **b**, $\rho(\tau)$. The initial model is described in the Methods.

¹School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287-1404, USA.

clearly in the $\phi(\tau)$ response between periods of 10^3 and 10^5 s. Assuming a 2% error in the $\phi(\tau)$ data in the 10^3 – 10^5 s range, a 1-km-thick layer with conductivity greater than 1.0 S m^{-1} (≥ 2.25 log units conductivity above the conductivity of the upper mantle at 409-km depth) could be detectable by MT/GDS methods (in agreement with χ^2 arguments, see Supplementary Information). Performing this analysis for high-conductivity layers of differing thickness gives an effective minimum conductance of $\sim 1,000 \text{ S}$ for a detectable high-conductivity layer near 410-km depth. The effective minimum conductance is dependent on the surrounding mantle conductivity structure and may vary regionally (see Supplementary Information).

In the transition-zone water filter hypothesis¹, a 410-km-deep melt layer may be composed of mafic material and contain high concentrations of incompatible elements. The melt may be hydrated and denser than the surrounding upper mantle, allowing the formation of a stable melt layer atop the 410-km-deep seismic discontinuity. Because the composition of a 410-km-deep melt is uncertain, the electrical conductivities of various melts are considered (Fig. 2). Electrical conductivity of silicate melts decreases with increasing pressure up to 2 GPa; at greater pressure, the dependence is reduced⁹. The electrical conductivity of basaltic melt at $1,420^\circ\text{C}$ and 2.5 GPa is $\sim 6.0 \text{ S m}^{-1}$, whereas for rhyolitic melt it is $\sim 2.0 \text{ S m}^{-1}$ (ref. 8). Alternative compositions such as (Mg,Fe)O fluids may be possible in this depth range, but experimental verification is lacking. Supercritical H_2O may also be possible; shock experiments yield electrical conductivity of $\sim 100 \text{ S m}^{-1}$ at 15 GPa and $\sim 1,000^\circ\text{C}$ (ref. 17).

At pressure–temperature conditions near 410-km depth, mafic melt is denser than olivine^{18–20}. It is not known whether fluids/melts of (Mg,Fe)O compositions would be gravitationally stable at these conditions. Shock experiments indicate that at the pressure–temperature conditions near 410-km depth, supercritical H_2O may be less dense than olivine¹⁷ and unable to form a stable layer. H_2O added to a silicate melt decreases its density²¹ and may increase its electrical conductivity⁷. A mafic melt at 410-km depth will remain denser than the surrounding upper mantle if it contains less than $\sim 6.0 \text{ wt\% H}_2\text{O}$ (refs 21, 22). The electrical conductivity enhancement for 6.0 wt% H_2O in a rhyolitic melt is ~ 0.25 log units above that of a dry rhyolitic melt⁷. The effects of water on the electrical conductivity of mafic

melts are unknown. We use an electrical conductivity of $\sim 6.0 \text{ S m}^{-1}$ as that of a melt near 410-km depth in calculations of layer thickness.

To allow upper-mantle melting, upwelling material (from the transition zone) must be sufficiently hydrated. Assuming that the H_2O storage capacity of the upper mantle is controlled by olivine alone, solubility measurements indicate a storage capacity of $\sim 0.4 \text{ wt\% H}_2\text{O}$ ($6.5 \times 10^4 \text{ H per } 10^6 \text{ Si}$, equivalent to 3.27 mol% H_2O) at $1,100^\circ\text{C}$ and 12 GPa (refs 23, 24; see Supplementary Information). Accounting for H_2O partitioning between olivine and other mantle minerals, upper-mantle H_2O storage capacity could be between 0.4 and 0.55 wt% H_2O at 410-km depth²⁵. Upper-mantle melting may occur only when the transition zone contains more water than the upper mantle can stably sequester; 0.4 wt% H_2O is chosen as the minimum amount of water needed in the transition zone to induce upper-mantle melting. The maximum solubility of H_2O in transition-zone minerals is $\sim 3.0 \text{ wt\%}$ at similar pressure–temperature conditions²³, indicating that the transition zone can accommodate sufficient H_2O to supersaturate the upper mantle through upwelling events.

The southern Basin and Range region near Tucson, Arizona, is characterized by slow seismic wave speeds in the upper mantle (the upper 100–200 km) and fast wave speeds in the transition zone (500–900 km depth)²⁶. A low-velocity zone has been seismically imaged in the western US, for which a 20–90-km-thick partial melt layer was inferred at the base of the upper mantle (410 km depth)³. Physical observations of the southern Basin and Range suggest lateral heterogeneities in the shallow lithosphere and a 1D upper mantle and transition zone. To determine if the Tucson MT/GDS data set is sufficiently 1D, the D^+ inversion²⁷ is used^{2,27}. The D^+ inversion, while not a physical model, will yield the minimum χ^2 misfit for a 1D solution²⁷. If the χ^2 misfit of the D^+ inversion is nearly equal to the degrees of freedom (d.f.) of the data set, then the data are sufficiently 1D in nature. The D^+ models of the Tucson data yield $\chi^2 = 62$ (d.f. = 48), indicating that the data set is nearly 1D. The smooth inversion of the data of ref. 2 yields $\chi^2 = 74$. The melt-free regional conductivity versus depth model includes 0.23 wt% H_2O in upper-mantle olivine (to 410-km depth) and 0.6 wt% H_2O in the transition zone, yielding $\chi^2 = 128$. Addition of a high-conductivity layer at 410-km depth is warranted because the transition-zone water content in this model is greater than 0.4 wt% H_2O . The water content estimate for the transition zone depends on the transition-zone geotherm. The geotherm used in ref. 14 is 50 – 75°C hotter than that of ref. 16 in the transition zone. This hotter geotherm reduces the transition-zone water content to 0.40 wt% H_2O for the Tucson data, which reaches the threshold for requiring a melt zone at 410-km depth. Minimizing χ^2 in the model including melt results in a melt layer with a conductance of $\sim 3.0 \times 10^4 \text{ S}$, and improves the χ^2 to 70 (Fig. 3). A melt layer of pure mafic melt ($\sigma \approx 6.0 \text{ S m}^{-1}$) with a conductance of $\sim 3.0 \times 10^4 \text{ S}$ would be ~ 5 -km thick. Considering resolution constraints (see Supplementary Information), we estimate that the melt (or partial melt) layer in this region is 5–30 km thick, in agreement with the teleseismic study of Song *et al.*³.

Similar analyses of other regional data sets^{10–13} confirm that the mineralogy¹⁴ and geotherm¹⁶ used are in good agreement with field data. The regional conductivities are consistent with 0.003 and 0.010 wt% H_2O (upper mantle and transition zone, respectively) in the French Alps¹⁰, 0.12 and 0.20 wt% H_2O in the North Pacific¹¹, 0.11 and 0.15 wt% H_2O for the European average¹², and 0.02 and 0.15 wt% H_2O in the Canadian Shield¹³. These values are not consistent with a transition-zone water content greater than 0.4 wt% H_2O , therefore melt layers in these regions are not warranted. Disregarding water solubility arguments and adding high-conductivity layers into these regional models resulted in no significant reduction of χ^2 and in most cases required a significant reduction of the surrounding mantle conductivity (compared to the melt-free regional model) to maintain the model fit to the data

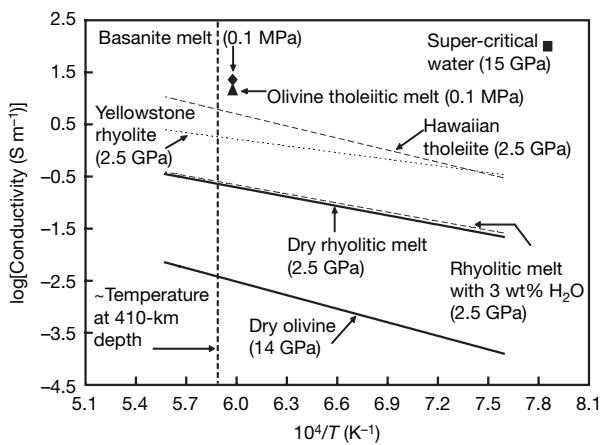


Figure 2 | Electrical conductivity of various melts as a function of inverse temperature. Upper solid line, rhyolitic melt at 2.5 GPa (ref. 7); nearby dashed line, rhyolitic melt with 3.0 wt% H_2O at 2.5 GPa (ref. 7); short-dashed line, Yellowstone rhyolite obsidian⁹ at 2.5 GPa; long-dashed line, Hawaiian tholeiite⁹ at 2.5 GPa. Square, supercritical water at $1,000^\circ\text{C}$ and 15 GPa (ref. 17). Triangle and diamond represent olivine tholeiitic melt⁹ and basanite⁸ at 0.1 MPa and $\sim 1,400^\circ\text{C}$, respectively. Also plotted is electrical conductivity of isotropic olivine¹⁴ for comparison (lower solid line). Vertical dashed line indicates approximate temperature at 410-km depth ($\sim 1,420^\circ\text{C}$).

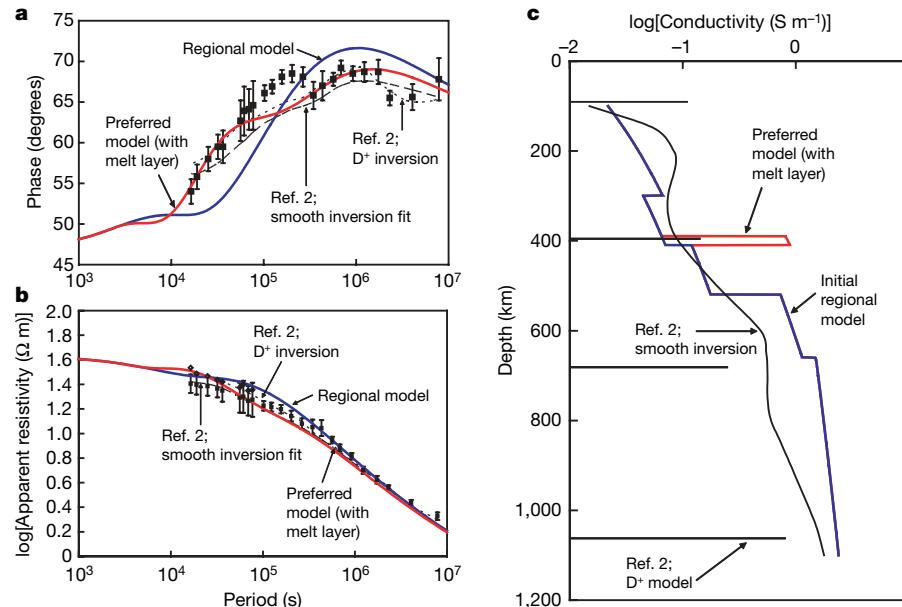


Figure 3 | Forward modelling results for the southern Basin and Range (Tucson) data set². The smoothed inversion of ref. 2 is the long-dashed line; the short-dashed line is the D^+ inversion². The red line is the preferred model with melt layer; the blue line is the melt-free ‘regional’ model. **a**, $\phi(\tau)$

models, and **b**, $\rho(\tau)$ models with the original data (black squares) and calculated upward biased $\rho(\tau)$ data (diamonds)². Error bars are the modified error bars of ref. 2, corresponding to $\pm 1\sigma$ except where noted. **c**, Electrical conductivity depth profiles.

(Supplementary Information and Supplementary Fig. 1). The requirement to reduce the electrical conductivity of the mantle above and below the high-conductivity layer indicates that a high-conductivity zone is not a robust feature of these regions. The models best fitting the Tucson² data did not require a reduction in the electrical conductivity of the surrounding mantle, suggesting that a high-conductivity layer at 410-km depth is a robust feature of the data.

1D forward modelling of MT/GDS responses of electrical conductivity depth profiles created using seismic and mineral physics constraints indicates that electrical methods can be used as imaging tools to detect thin conductive zones at 410-km depth. The estimated minimum detectable conductance for a high-conductivity layer at 410-km depth is $\sim 1,000$ S; however, this may vary regionally, depending on the surrounding electrical conductivity structure. The agreement between forward models and southern Basin and Range (Tucson)² MT/GDS data is enhanced when a melt layer at 410-km depth is added to the hydrogen-containing mineralogically based regional electrical conductivity depth profile, and is consistent with petrologic constraints on hydrogen-enhanced melting at these depths. This is the only region studied where this is the case. A melt (or partial melt) layer with conductance of $\sim 3.0 \times 10^4$ S (5–30 km thick) is in agreement with seismic studies³. These results do not support the global presence of a transition-zone water filter, but instead indicate its regional occurrence. Additional long-period MT/GDS field studies and stricter limits on melt composition and conductivity are needed to constrain the existence and properties of a 410-km-deep melt layer.

METHODS

The mineralogy of all models is based on that of ref. 14 with 60 vol.% olivine, 25 vol.% orthopyroxene and 15 vol.% clinopyroxene between 100-km and 300-km depth. Between 300-km and 410-km depth, the profile consists of 60 vol.% olivine and 40 vol.% clinopyroxene. The upper transition zone (410–520 km depth) is composed of 60 vol.% wadsleyite and 40 vol.% clinopyroxene. The lower transition zone (520–660 km depth) is composed of 60 vol.% ringwoodite and 40 vol.% akimotoite plus garnet. The lower mantle is composed of 75 vol.% Al-bearing perovskite (~ 2.9 wt% Al) and 25 vol.% magnesiowüstite. The geotherm of ref. 16 is used as the temperature profile for all models. A regional conductivity profile is created by adding water to olivine⁵ in the upper mantle and to wadsleyite and ringwoodite⁶ in the transition zone to better fit specific regional MT/GDS field

responses (that is, the melt-free regional model, blue line in Fig. 3). If the response indicates transition-zone conductivity consistent with transition-zone minerals having ≥ 0.4 wt% H₂O, then a high-conductivity (melt) layer is added to the model (red line in Fig. 3).

Received 25 January; accepted 10 May 2007.

1. Bercovici, D. & Karato, S. Whole-mantle convection and the transition-zone water filter. *Nature* **425**, 39–44 (2003).
2. Egbert, G. D., Booker, J. R. & Schultz, A. Very long period magnetotellurics at Tucson Observatory: Estimation of impedances. *J. Geophys. Res. B* **97**, 15113–15128 (1992).
3. Song, T. R. A., Helmberger, D. V. & Grand, S. P. Low-velocity zone atop the 410-km seismic discontinuity in the northwestern United States. *Nature* **427**, 530–533 (2004).
4. Revenaugh, J. & Sipkin, S. A. Seismic evidence for silicate melt atop the 410-km mantle discontinuity. *Nature* **369**, 474–476 (1994).
5. Wang, D., Mookherjee, M., Xu, Y. S. & Karato, S. The effect of hydrogen on the electrical conductivity in olivine. *Nature* **443**, 977–980 (2006).
6. Huang, X. G., Xu, Y. S. & Karato, S. I. Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite. *Nature* **434**, 746–749 (2005).
7. Gaillard, F. Laboratory measurements of electrical conductivity of hydrous and dry silicic melts under pressure. *Earth Planet. Sci. Lett.* **218**, 215–228 (2004).
8. Tyburczy, J. A. & Fisler, D. K. in *Mineral Physics and Crystallography: A Handbook of Physical Constants* (ed. Ahrens, T. J.) 185–208 (American Geophysical Union, Washington DC, 1995).
9. Tyburczy, J. A. & Waff, H. S. Electrical conductivity of molten basalt and andesite to 25 kilobars pressure — Geophysical significance and implications for charge transport and melt structure. *J. Geophys. Res.* **88**, 2413–2430 (1983).
10. Tarits, P., Hautot, S. & Perrier, F. Water in the mantle: Results from electrical conductivity beneath the French Alps. *Geophys. Res. Lett.* **31**, L06612, doi:10.1029/2003GL019227 (2004).
11. Lizarralde, D., Chave, A., Hirth, G. & Schultz, A. Northeastern Pacific mantle conductivity profile from long-period magnetotelluric sounding using Hawaii-to-California submarine cable data. *J. Geophys. Res. B* **100**, 17837–17854 (1995).
12. Olsen, N. Long-period (30 days–1 year) electromagnetic sounding and the electrical conductivity of the lower mantle beneath Europe. *Geophys. J. Int.* **138**, 179–187 (1999).
13. Schultz, A., Kurtz, R. D., Chave, A. D. & Jones, A. G. Conductivity discontinuities in the upper mantle beneath a stable craton. *Geophys. Res. Lett.* **20**, 2941–2944 (1993).
14. Xu, Y. S., Shankland, T. J. & Poe, B. T. Laboratory-based electrical conductivity in the Earth’s mantle. *J. Geophys. Res. B* **105**, 27865–27875 (2000).
15. Simpson, F. Resistance to mantle flow inferred from electromagnetic strike of the Australian upper mantle. *Nature* **412**, 632–635 (2001).

16. Brown, J. M. & Shankland, T. J. Thermodynamic parameters in the Earth as determined from seismic profiles. *Geophys. J. R. Astron. Soc.* **66**, 579–596 (1981).
17. Mitchell, A. C. & Nellis, W. J. Equation of state and electrical conductivity of water and ammonia shocked to 100 GPa (1Mbar) pressure range. *J. Chem. Phys.* **76**, 6273–6281 (1982).
18. Ohtani, E. & Maeda, M. Density of basaltic melt at high pressure and stability of the melt at the base of the lower mantle. *Earth Planet. Sci. Lett.* **193**, 69–75 (2001).
19. Ohtani, E., Nagata, Y., Suzuki, A. & Kato, T. Melting relations of peridotite and the density crossover in planetary mantles. *Chem. Geol.* **120**, 207–221 (1995).
20. Suzuki, A., Ohtani, E. & Kato, T. Density and thermal expansion of a peridotite melt at high pressure. *Phys. Earth Planet. Inter.* **107**, 53–61 (1998).
21. Matsukage, K. N., Jing, Z. C. & Karato, S. Density of hydrous silicate melt at the conditions of Earth's deep upper mantle. *Nature* **438**, 488–491 (2005).
22. Sakamaki, T., Suzuki, A. & Ohtani, E. Stability of hydrous melt at the base of the Earth's upper mantle. *Nature* **439**, 192–194 (2006).
23. Kohlstedt, D. L., Keppler, H. & Rubie, D. C. Solubility of water in the α , β and γ phases of $(\text{Mg},\text{Fe})_2\text{SiO}_4$. *Contrib. Mineral. Petrol.* **123**, 345–357 (1996).
24. Bell, D. R., Rossman, G. R., Maldener, J., Endisch, D. & Rauch, F. Hydroxide in olivine: A quantitative determination of the absolute amount and calibration of the IR spectrum. *J. Geophys. Res.* **108**, B2105, doi:10.1029/2001JB000679 (2003).
25. Hirschmann, M. M., Aubaud, C. & Withers, A. C. Storage capacity of H_2O in nominally anhydrous minerals in the upper mantle. *Earth Planet. Sci. Lett.* **236**, 167–181 (2005).
26. van der Lee, S. & Nolet, G. Upper mantle S velocity structure of North America. *J. Geophys. Res. B* **102**, 22815–22838 (1997).
27. Parker, R. L. The inverse problem of electromagnetic induction — Existence and construction of solutions based on incomplete data. *J. Geophys. Res.* **85**, 4421–4428 (1980).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Constable and E. Garnero for comments and discussion. This work was supported by the National Science Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A.T. (jim.tyburczy@asu.edu).

LETTERS

Recuperation of nitrogen cycling in Amazonian forests following agricultural abandonment

Eric A. Davidson¹, Cláudio J. Reis de Carvalho², Adelaine Michela Figueira³, Françoise Yoko Ishida³, Jean Pierre H. B. Ometto³, Gabriela B. Nardoto³, Renata Tuma Sabá², Sanae N. Hayashi⁴, Eliane C. Leal⁴, Ima Célia G. Vieira⁴ & Luiz A. Martinelli³

Phosphorus (P) is generally considered the most common limiting nutrient for productivity of mature tropical lowland forests growing on highly weathered soils^{1–5}. It is often assumed that P limitation also applies to young tropical forests, but nitrogen (N) losses during land-use change may alter the stoichiometric balance of nutrient cycling processes. In the Amazon basin, about 16% of the original forest area has been cleared⁶, and about 30–50% of cleared land is estimated now to be in some stage of secondary forest succession following agricultural abandonment⁷. Here we use forest age chronosequences to demonstrate that young successional forests growing after agricultural abandonment on highly weathered lowland tropical soils exhibit conservative N-cycling properties much like those of N-limited forests on younger soils in temperate latitudes. As secondary succession progresses, N-cycling properties recover and the dominance of a conservative P cycle typical of mature lowland tropical forests re-emerges. These successional shifts in N:P cycling ratios with forest age provide a mechanistic explanation for initially lower and then gradually increasing soil emissions of the greenhouse gas nitrous oxide (N_2O). The patterns of N and P cycling during secondary forest succession, demonstrated here over decadal timescales, are similar to N- and P-cycling patterns during primary succession as soils age over thousands and millions of years, thus revealing that N availability in terrestrial ecosystems is ephemeral and can be disrupted by either natural or anthropogenic disturbances at several timescales.

Ecologists have long noted that tropical forests growing on highly weathered soils exhibit conservative P-cycling processes, whereas conservative N-cycling properties are more common on younger soils, including most temperate forests and montane forests^{1,2}. This pattern was demonstrated along a soil age chronosequence in the Hawaiian Islands³, where N, which is derived primarily from the atmosphere, is in short supply in the youngest volcanic soils and gradually accumulates as soils age. In contrast, rock-derived P is more abundant in young soils but becomes bound in unavailable forms to soil minerals as soil weathering proceeds over thousands and millions of years. In global-scale analyses, the N:P ratios of green foliage⁴ and litterfall⁵ of mature forests have been shown to increase with decreasing latitude, indicating generally increasing P conservation and decreasing N conservation with soil age.

Although these stoichiometric generalizations seem robust for mature forests, accelerating land-use change is altering tropical landscapes worldwide, and the consequences for nutrient cycling in secondary forests are unclear⁸. Secondary tropical forests are playing an increasingly important part in maintaining genetic diversity⁹ and hydrological functioning of altered landscapes¹⁰, but biogeochemical processes remain poorly studied in tropical secondary forest succession.

Forest clearing causes an initial loss of nutrients from Amazonian terrestrial ecosystems through fire, erosion, soil emissions of gases, harvesting of timber and hydrologic leaching of nutrients^{8,11}. Additional losses occur as cattle or crops are harvested and as fire is used as a management tool to prepare fields for planting and to control pasture weeds¹². Both N and P can be lost as particulates during biomass burning¹³, but, in contrast to P, N is also volatilized as a gas. Nitrate also generally leaches from soils more readily than does phosphate. As a consequence of these N losses, net N mineralization, net nitrification, nitrate leaching and soil efflux of N_2O often decline as tropical cattle pastures age^{14–17}. Fertilization can maintain agricultural productivity, but where fertilization is not economically viable, the cleared land is often abandoned and a secondary forest begins to grow. Aggrading forests create a strong demand for essential plant nutrients. The objective of this study was to use space-for-time substitutions in secondary forest age chronosequences to describe patterns in the indicators of N and P nutrient cycling during tropical secondary forest succession.

Three forest-age chronosequences, including stands ranging in age from 3 to 70 yr and remnant mature forests, were established in the Brazilian state of Pará, in eastern Amazonia. A complete set of seven indicators of N-cycling rates was obtained for the chronosequence in our main study site in the municipality of São Francisco do Pará. To provide true replication, a second chronosequence was established in the municipality of Capitão Poço, and a third chronosequence was constructed from previously published studies on a ranch in the municipality of Paragominas. The São Francisco do Pará and Paragominas municipalities are about 200 km apart, with Capitão Poço roughly in the middle. Not all of the indicators could be measured in the second and third chronosequence, but each indicator is represented by at least two replicate chronosequences. Soil texture varied among the chronosequences, but was relatively uniform within each chronosequence (Supplementary Table 1). The dominant vegetation of the region was once moist lowland tropical forest, but is now a mosaic of secondary forests, agricultural fields, cattle pastures and tree crops¹⁸ (see Supplementary Information for more site information).

All of the indicators derived from analysis of green foliage, litterfall, soil and trace gas emissions are consistent with a conservative N cycle in the young successional forests, recovery of N-cycling processes as succession proceeds, and a leaky N cycle in advanced stages of secondary succession and in mature forests (Fig. 1). The log-linear relationships indicate that the largest changes in N-cycling indicators occur early during succession and that the rate of change declines as the secondary forests mature.

The first indicator, foliar ^{15}N , increases with increasing forest age in both São Francisco do Pará and Capitão Poço chronosequences

¹The Woods Hole Research Center, 149 Woods Hole Road, Falmouth, Massachusetts 02540-1644, USA. ²EMBRAPA Amazônia Oriental, C. P. 48, Belém, PA 66.095-100, Brazil. ³CENA, University of São Paulo, Avenue Centenário, 303, Piracicaba, SP 13.416-000, Brazil. ⁴Departamento de Botânica, Museu Paraense Emílio Goeldi, Belém, PA 66.040-179, Brazil.

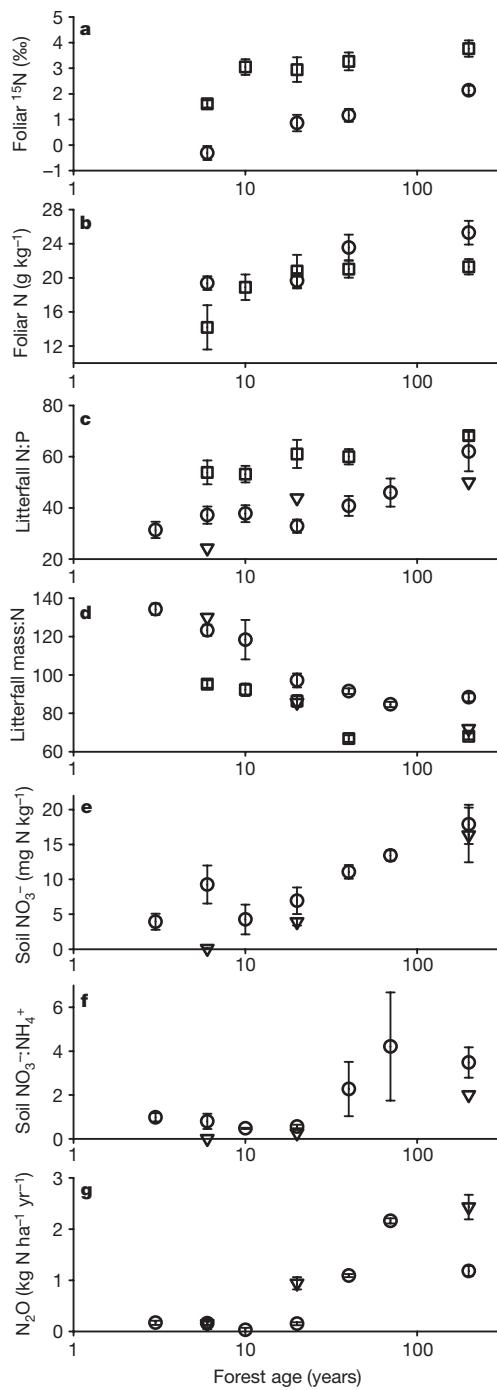


Figure 1 | Indicators of N and P cycling along secondary successional forest chronosequences. a–g, Foliar delta ^{15}N (a); foliar N concentration (b); litterfall N:P ratios (c); litterfall mass-to-N ratios (d); soil nitrate (e); soil nitrate:ammonium ratios (f); and annual soil emissions of nitrous oxide (g). Error bars indicate s.e.m. within each forest age at each chronosequence location. The chronosequences are represented as circles for São Francisco do Pará, squares for Capitão Poço and triangles for Paragominas. The effect of forest age, either as a ranking factor or as the logarithm of age (assuming an age of 200 yr for the mature forest; see Methods) is significant in analysis of covariance (ANCOVA) for all seven indices ($P < 0.05$; see Supplementary Table 3 for all P values). The site-effect is significant for foliar ^{15}N , litterfall N:P, litterfall mass:N and N_2O . Results are nearly identical when the mature forests (of unknown age) are omitted from the analysis, except that the age-effect is not significant for N_2O and is only marginally significant for the nitrate:ammonium ratio ($P = 0.119$ and $P = 0.057$, respectively).

(Fig. 1a). The foliar ^{15}N values of the mature forests of these sites are within the range commonly measured in mature tropical forest foliage ($3.7\text{‰} \pm 3.5$; error of one standard deviation¹⁹), but the values in the youngest forest at São Francisco do Pará (-0.5‰) are as low as those usually measured in temperate forest foliage ($-2.8\text{‰} \pm 2.0$; error of one standard deviation¹⁹). Enriched foliar ^{15}N in the older forests is indicative of a leaky N cycle, in which isotopically light N is lost from the ecosystem owing to fractionation during nitrification and denitrification, leaving isotopically enriched N behind²⁰. In contrast, little N is lost in the conservative N cycle of young secondary forests, resulting in little N fractionation or ^{15}N enrichment. Increasing concentrations of fresh foliar N concentrations with increasing forest age (Fig. 1b), provide a second indicator of increasing N availability as secondary forest succession progresses.

Third, the N:P ratios of fine litterfall increased with forest age (Fig. 1c) for all three chronosequences. The mean litterfall N:P ratios in the mature forests (50–68) bound the global mean of 62 for tropical forests⁵. In contrast, the mean litterfall N:P values of 31 and 24 in the youngest forests of the São Francisco do Pará and Paragominas chronosequences are closer to the global mean of 29 for broadleaf temperate forests⁵. The mean N:P ratio in fresh foliage ranged from 27 to 35 and did not vary systematically across forest ages. Hence, only in the advanced successional and mature forests did the trees reabsorb more P relative to N before leaf abscission. Fourth, the litterfall mass:N ratio also declined with forest age (Fig. 1d). The values in the young forests (100–140) are similar to N-limited temperate forests, whereas the advanced and mature forests values (68–88) are common for mature lowland tropical forests¹. Although both foliar N and P may be reabsorbed before leaf fall, the balance shifts from N economy to P economy in litterfall during secondary forest succession.

Fifth, extractable soil nitrate increased with forest age (Fig. 1e), indicating increasing availability of soil N as the forests mature. Ammonium (which tended to decrease with forest age, but not significantly so) is often the dominant form of inorganic soil N in N-limited systems, whereas nitrate accumulates where available N is more abundant^{14,21}. Hence, in a related sixth index, the ratio of extractable nitrate to extractable ammonium increased from values ≤ 1 for forests ≤ 20 yr to values > 1 for forests ≥ 40 yr (Fig. 1f).

Seventh, although the more clay-rich Paragominas site exhibited higher soil emissions of N_2O than the sandy São Francisco do Pará site, N_2O emissions increased with forest age at both sites (Fig. 1g). Higher rates of nitrate leaching in mature forests compared to young forests have also been measured at the Paragominas site²². These differences in N losses across forest ages reflect current biogeochemical fluxes and are consistent with the time-integrated indicator of ecosystem N loss and fractionation provided by foliar ^{15}N analyses²⁰, which reflects decades of previous slash-and-burn management in the young forests and decades of regrowth in the older forests.

These results have important implications for rates of regrowth of secondary forest and the sustainability of slash-and-burn agriculture, which depends on the accumulation of nutrients in fallow vegetation as the source of nutrients for the next cycle of slash-and-burn agriculture. Rates of secondary forest growth in Amazonia have been inversely correlated with the number of fires during the agricultural phases²³. A large fraction of biomass N is often lost during fires¹³, depleting the pool of actively cycling ecosystem N and provoking a N limitation and a conservative N cycle after repeated fire⁸. The time needed between cycles to re-accumulate needed nutrients for the next agricultural phase increases when the pool of available nutrients, such as available N, declines with each fire cycle. Further evidence for nutrient limitation comes from nutrient amendment experiments in which biomass accumulation in young Amazonian secondary forests responded to N and not P additions in one study²⁴ and showed species-specific responses in another study²⁵.

Although lower emissions of N_2O in secondary compared to mature tropical forests have been previously observed^{14,17,26,27}, here

we show that increasing emissions of N_2O as the successional forests age can be understood in terms of gradual recuperation of several N-cycling processes. A legacy of large N losses during an agricultural phase would probably result in slow rates of recovery of N cycling⁸ and slow rates of increases of N_2O emissions during secondary succession. Conversely, acceleration of N-cycle recuperation and rapid increases in N_2O emissions could result from the emergence of a dominant N-fixing species during secondary forest succession²⁷.

Recent isotopic evidence suggests that biological N fixation may contribute significantly to recuperation of the N cycle during the first 25 yr of secondary succession of Amazonian forests²⁸, although quantitative estimates are poorly constrained. In our study, however, no difference in foliar ^{15}N enrichments between legume and non-legume species was observed (Supplementary Table 2), although leguminous species had higher foliar N concentrations and lower C:N ratios. Leguminous tree species in closed-canopy tropical forests often exhibit N-rich tissues without necessarily fixing significant amounts of nitrogen²⁹. We can neither rule out nor support the importance of N fixation in this study. Atmospheric deposition inputs of nitrogen in this region (2–6 kg ha⁻¹ yr⁻¹; ref. 22) are in the same range of 3–8 kg ha⁻¹ yr⁻¹ that accumulates in woody biomass of these successional forests, but are not enough to account for about 11 kg ha⁻¹ yr⁻¹ that accumulates in foliar biomass during canopy development^{22,24}. The top 10 cm of mineral soil contains ample organic-N stocks (≥ 1000 kg ha⁻¹ of N atoms) that could supply the regrowing forest if a small fraction is gradually mineralized to a bioavailable form, and considerably more is present at lower soil depths²². The rate of recuperation of N-cycling processes during secondary succession may reflect, in part, the kinetics of mobilization of recalcitrant forms of soil N to an actively cycling N pool, as well as the legacy of the degree of degradation during agricultural phases.

The patterns of N and P cycling shown here for secondary succession parallel those previously demonstrated for primary succession. Actively cycling N in terrestrial ecosystems can be lost either by land-use change, such as forest clearing, burning and agricultural practices, or by natural processes such as fires, landslides, glaciers and volcanic activity. Just as accumulation of total ecosystem N alleviates an N limitation as soils age over thousands and millions of years, actively cycling N accumulates over decades and centuries during secondary forest succession, resulting in a similar successional trajectory from a conservative N cycle following agricultural abandonment to the leaky-N and conservative-P cycles expected in mature lowland tropical forests on old soils.

METHODS SUMMARY

In the São Francisco do Pará municipality, two farms were selected to represent each forest age of 3, 6, 10, 20, 40 and 70 yr. In the Capitão Poço municipality, one farm was identified for each forest age of 6, 10, 20 and 40 yr. Two 20 m × 20 m plots were established on each farm. These forests were regrowing after abandonment of crop fields. Two plots were also established in each remnant mature forest at São Francisco do Pará and Capitão Poço. Measurements were made from October 2000 through to June 2002 in São Francisco do Pará, and from February 2004 to January 2005 at Capitão Poço. For the Paragominas chronosequence, data from previous studies^{17,22,24} were assembled to construct a chronosequence that included 6- and 20-yr-old forests on abandoned cattle pastures and a mature forest, all on the same soil type and ranch. An ANCOVA was used to test the effects of site (São Francisco do Pará, Capitão Poço and Paragominas) and forest age on mean values of each N and P indicator for each site and age. Because of uncertainty of the ages of the mature forests, the effects of forest age were analysed three ways. First, age was converted to a rank score from 1 to 7 (3-, 6-, 10-, 20-, 40- and 70-yr-old secondary forests and the mature forest, respectively). Second, the ages of the mature forests were conservatively estimated at 200 yr, and the logarithm of forest age was used as a continuous variable. Third, this analysis was repeated with the data from mature forests excluded to test the effects of log-age only in the successional forests of known age. The minor differences in results of these statistical tests are presented in the Fig. 1 legend and in Supplementary Table 3.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 April; accepted 2 May 2007.

- Vitousek, P. M. Litterfall, nutrient cycling, and nutrient limitations in tropical forests. *Ecology* 65, 285–298 (1984).
- Walker, T. W. & Syers, J. K. The fate of phosphorus during pedogenesis. *Geoderma* 15, 1–19 (1976).
- Hedin, L. O., Vitousek, P. M. & Matson, P. A. Nutrient losses over four million years of tropical forest development. *Ecology* 84, 2231–2255 (2003).
- Reich, P. B. & Okelesyn, J. Global patterns of plant leaf N and P in relation to temperature and latitude. *Proc. Natl Acad. Sci.* 101, 11001–11006 (2004).
- McGroddy, M. E., Daufresne, T. & Hedin, L. Scaling of C:N:P stoichiometry in forests worldwide: implications of terrestrial redfield-type ratios. *Ecology* 85, 2390–2401 (2004).
- INPE (Instituto Nacional de Pesquisas Espaciais) *Monitoramento Ambiental da Amazônia por Satélite* (http://www.obt.inpe.br/prodes/apresentacao_prodes.ppt) (2004).
- Hirsch, A. I., Little, W. S., Houghton, R. A., Scott, N. A. & White, J. D. The net carbon flux due to deforestation and forest re-growth in the Brazilian Amazon: analysis using a process-based model. *Glob. Change Biol.* 10, 908–924 (2004).
- McGrath, D. A., Smith, C. K., Ghosh, H. L. & de Assis Oliveira, F. Effects of land-use change on soil nutrient dynamics in Amazônia. *Ecosystems* 4, 625–645 (2001).
- Vieira, I. C. G., de Paiva Salomão, R., de Araujo Rosa, N., Nepstad, D. C. & Roma, J. C. Renascimento da floresta no rastro da agricultura. (The rebirth of the forest in the wake of agriculture.) *Ciência Hoje* 20, 38–45 (1996).
- Nepstad, D. C., Moutinho, P. & Markewitz, D. In *Biogeochemistry of the Amazon Basin and its Role in a Changing World* (eds McClain, M. E., Victoria, R. L. & Richey, J. E.) 139–155 (Oxford Univ. Press, New York, 2001).
- Davidson, E. A. et al. In *Ecosystems and Land Use Change*. Geophysical Monograph Series 153 (eds DeFries, R., Asner, G. & Houghton R.) 147–158 (American Geophysical Union, Washington DC, 2004a).
- Dias-Filho, M., Davidson, E. A. & Carvalho, C. In *The Biogeochemistry of the Amazon Basin* (eds McClain, M. E., Victoria, R. L. & Richey, J. E.) 84–105 (Oxford Univ. Press, New York, 2001).
- Kaufman, J. B., Cummings, D. L., Ward, D. E. & Babbitt, R. Fire in the Brazilian Amazon: biomass, nutrient pools, and losses in slashed primary forests. *Oecologia* 104, 397–409 (1995).
- Davidson, E. A., Keller, M., Erickson, H. E., Verchot, L. V. & Veldkamp, E. Testing a conceptual model of soil emissions of nitrous and nitric oxides. *Bioscience* 50, 667–680 (2000).
- Keller, M., Veldkamp, E., Weitz, A. M. & Reiners, W. A. Effect of pasture age on soil trace-gas emissions from a deforested area of Costa Rica. *Nature* 365, 244–246 (1993).
- Melillo, J. M. et al. Nitrous oxide emissions from forests and pastures of various ages in the Brazilian Amazon. *J. Geophys. Res.* 106, 34,179–34,188 (2001).
- Verchot, L. V. et al. Land use change and biogeochemical controls of nitrogen oxide emissions from soils in eastern Amazonia. *Global Biogeochem. Cycles* 13, 31–46 (1999).
- Vieira, I. C. G. et al. Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazonia. *Remote Sens. Environ.* 87, 470–481 (2003).
- Martinelli, L. A. et al. Nitrogen stable isotopic composition of leaves and soil: Tropical versus temperate forests. *Biogeochemistry* 46, 45–65 (1999).
- Amundson, R. et al. Global patterns of the isotopic composition of soil and plant nitrogen. *Global Biogeochem. Cycles* 17, 1031 (2003).
- Vitousek, P. M., Gosz, J. R., Grier, C. C., Melillo, J. M. & Reiners, W. A. A comparative analysis of potential nitrification and nitrate mobility in forest ecosystems. *Ecol. Monogr.* 52, 155–177 (1982).
- Markewitz, D., Davidson, E. A., Moutinho, P. & Nepstad, D. C. Nutrient loss and redistribution after forest clearing on a highly weathered soil in Amazonia. *Ecol. Appl.* 14, S177–S199 (2004).
- Zarin, D. J. et al. Legacy of fire slows carbon accumulation in Amazonian forest regrowth. *Front. Ecol. Environ.* 3, 365–369 (2005).
- Davidson, E. A. et al. Nitrogen and phosphorus limitation of biomass growth in a tropical secondary forest. *Ecol. Appl.* 14, S150–S163 (2004b).
- Gehring, C., Denich, M., Kanashiro, M. & Vlek, P. L. G. Response of secondary vegetation in Eastern Amazonia to relaxed nutrient availability constraints. *Biogeochemistry* 45, 223–241 (1999).
- Keller, M. & Reiners, W. A. Soil-atmosphere exchange of nitrous oxide, nitric oxide, and methane under secondary succession of pasture to forest in the Atlantic lowlands of Costa Rica. *Global Biogeochem. Cycles* 8, 399–409 (1994).
- Erickson, H., Keller, M. & Davidson, E. Nitrogen oxide fluxes and nitrogen cycling during post-agricultural succession and forest fertilization in the humid tropics. *Ecosystems* 4, 67–84 (2001).
- Gehring, C. Biological nitrogen fixation in secondary regrowth and mature rainforest of central Amazonia. *Agr. Ecosyst. Environ.* 111, 237–252 (2005).

29. Vitousek, P. M. et al. Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry* **57**, 1–45 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. de F. R. Pantoja for litterfall collections, A. Almeida for assistance with establishing the chronosequences, R. Figueiredo for assistance with project and data management, and the Large-Scale Biosphere–Atmosphere (LBA) training and education programme for student stipends. This work was supported by grants from the LBA-Ecology programme of NASA.

Author Contributions E.A.D. established project design, led the fund-raising effort, supervised the field work of F.Y.I. and R.T.S., and wrote the paper. C.J.R.d.C.

conducted laboratory analyses of soils and litter. A.M.F. collected foliar samples at São Francisco and conducted isotopic analyses, under the supervision of J.P.H.B.O. and G.B.N. F.Y.I. led and R.T.S. assisted with field measurements of trace gas fluxes. S.N.H. conducted litterfall studies and E.C.L. collected soils and fresh foliage at Capitão Poço, both under the supervision of I.C.G.V. A.M.F., F.Y.I., J.P.H.B.O., G.B.N., and L.A.M. contributed to an early draft of the manuscript in Portuguese, and L.A.M. supervised the work of A.M.F. I.C.G.V. established the chronosequence study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.A.D. (edavidson@whrc.org).

METHODS

Site descriptions. The remnant mature forests in the municipalities of São Francisco do Pará and Capitão Poço are located at 1° 15' S, 47° 47' W and 02° 12' S, 47° 22' W, respectively, and the secondary forests are located in those vicinities. The Fazenda Vitoria ranch in the municipality of Paragominas is located at 2° 59' S, 47° 31' W. Mean annual precipitation is 2,200 mm for São Francisco do Pará and Capitão Poço and 1,800 mm for Paragominas. The dominant soils are Typic Hapludults in São Francisco do Pará and Capitão Poço, and Typic Hapustox at the Paragominas ranch (see Supplementary Table 1 for surface soil characteristics). Human settlement in this region expanded during the era of rubber extraction in the late nineteenth century. Government programmes for distributing rural lands for agricultural development, mostly by small land-holders, increased forest clearing in the mid-twentieth century. Many farms have now undergone nine or more cycles of slash-and-burn agriculture¹⁸. Cattle ranching and logging were the main forces of deforestation in the 1960–80s in Paragominas, including the ranch used in this study, which was established in 1969 (ref. 22).

Method details. Soil emissions of N₂O were measured at São Francisco do Pará and Paragominas using syringe sampling of static chambers and gas chromatography with an electron capture detector¹⁷. Three chamber fluxes were measured per date in each of the twenty-six plots at São Francisco do Pará, with five dates in each of the dry and wet seasons at São Francisco do Pará. Three collections (0.25 m² per collection) of fine litterfall were made monthly for a year in each plot at both São Francisco do Pará and Capitão Poço. Soil inorganic N was extracted in 1 M KCl from triplicate soil samples collected from the top 10 cm in each plot during the rainy season. Fresh foliar samples were collected in the 6-, 20- and 40-yr-old successional forests and the mature forest of São Francisco do Pará and Capitão Poço. At São Francisco do Pará, fully expanded leaves were collected for the dominant species at each site, according to the species importance values indices¹⁸. At Capitão Poço, all leaves within 1 m × 2 m miniplots were harvested and composited. Finely ground foliar samples were analysed for C and N concentrations using a Carlo-Erba CHN analyser, for C and N stable isotope ratios using a Delta Plus ThermoQuest-Finnigan mass spectrometer, and for P by acid digestion followed by colorimetric spectrophotometry.

LETTERS

The significance of nitrification for oceanic new production

Andrew Yool¹, Adrian P. Martin¹, Camila Fernández^{2,3} & Darren R. Clark⁴

The flux of organic material sinking to depth is a major control on the inventory of carbon in the ocean¹. To first order, the oceanic system is at equilibrium such that what goes down must come up². Because the export flux is difficult to measure directly, it is routinely estimated indirectly by quantifying the amount of phytoplankton growth, or primary production, fuelled by the upward flux of nitrate³. To do so it is necessary to take into account other sources of biologically available nitrogen. However, the generation of nitrate by nitrification in surface waters has only recently received attention. Here we perform the first synthesis of open-ocean measurements of the specific rate of surface nitrification^{4–12} and use these to configure a global biogeochemical model^{13,14} to quantify the global role of nitrification. We show that for much of the world ocean a substantial fraction of the nitrate taken up is generated through recent nitrification near the surface. At the global scale, nitrification accounts for about half of the nitrate consumed by growing phytoplankton. A consequence is that many previous attempts to quantify marine carbon export, particularly those based on inappropriate use of the *f*-ratio (a measure of the efficiency of the ‘biological pump’), are significant overestimates.

On timescales of hundreds to thousands of years, the ocean is the Earth’s largest active reservoir of carbon and is the dominant sink for anthropogenic carbon dioxide¹⁵. This is principally a result of the solubility of CO₂ and the volume of the ocean, but also of the ‘biological pump’¹. This ‘pump’ is driven primarily by biological processes that create particulate organic matter (POM), with primary production by phytoplankton dominating this either directly or indirectly. Although most POM is remineralized within the surface ocean as it sinks¹⁶, there is a sufficient export flux to increase concentrations of dissolved inorganic carbon at depth¹⁷. This sequestration of carbon means that accurate quantification of the biological pump is crucial for understanding the Earth system’s response to the continuing anthropogenic perturbation of the carbon cycle^{18,19}.

Quantifying the biological pump’s strength is not straightforward. Direct measurements require instruments to operate at demanding depths, sampling sinking material over periods of at least a month. The export flux is therefore often estimated indirectly by assessing the fraction of primary production that eventually constitutes export. This approach requires care, because phytoplankton use nutrients recycled from POM before it can sink out of the surface ocean, leading to the risk of double-counting.

The nitrogen cycle potentially offers a method of separating production driven by recycled nutrient from that leading to export. The inorganic nitrogen pool that supports production occurs largely as two nutrients, ammonium and nitrate, formed through distinct pathways. When nitrogen is metabolized through the foodweb it returns to the dissolved pool as ammonium. Nitrate is generated through the oxidation of ammonium, a biologically mediated

process known as nitrification. Until relatively recently, nitrification was believed to occur almost entirely at depth, possibly because of inhibition by light^{20,21}. The assumed depth separation of these pathways has been used to define ‘new’ (P_{new}) and ‘regenerated’ (P_{regen}) primary production³: the former is driven by nitrate that reaches surface waters primarily from depth, by means of vertical transport; the latter is fuelled by ammonium from recycling *in situ* in the euphotic zone.

At equilibrium, the upward flux of nitrate (plus smaller fluxes driven by nitrogen fixation and deposition) must balance the downward flux of sinking POM. Measuring nitrate use by phytoplankton should, therefore, provide an indirect estimate of the efficiency of the biological pump. This was formalized in a property known as the *f*-ratio², defined as $P_{\text{new}}/(P_{\text{new}} + P_{\text{regen}})$.

Classically, areas in which most primary production is fuelled by nitrate, such as the mesotrophic high latitudes, have a much higher *f*-ratio than areas such as the oligotrophic subtropical gyres, where most production is simply the product of recycled nutrient. As noted when it was first defined, the utility of the *f*-ratio as a proxy for export production depends crucially on the depth separation of primary production and nitrification. In the past decade, advances in measurement techniques and technologies have led to improvements in quantifying nitrification and to a greater understanding of its significance both for the nitrogen cycle in general and for ‘new’ production^{6–12,22–25}.

However, nitrification remains a difficult process to quantify, and direct observations for the open ocean are still scarce. Figure 1a represents the locations of extant observations (see Supplementary Information): nine data sets in all, comprising 431 measurements. Although sparse, observations have been taken in the Atlantic, Pacific and Southern oceans and in tropical, subtropical and subpolar areas.

For maximum consistency across data sets, we define the specific nitrification rate, λ_{nitrif} as the daily rate of ammonium oxidation divided by the corresponding ammonium concentration (with just one exception; see Supplementary Information). Figure 1b shows that λ_{nitrif} varies substantially, with rates spanning four orders of magnitude. There is no clear increase in λ_{nitrif} with depth, which would be indicative of inhibition by light^{20,21}. Furthermore, although two of the smaller studies used dark⁷ or deep⁶ incubations, almost all of the experiments were performed under *in situ* (or simulated) light conditions for periods of 6 h (ref. 4), 12 h (refs 8, 10) or 24 h (refs 9, 11, 12).

The mean of λ_{nitrif} over all observations is 0.550 d^{-1} . However, the distribution is not gaussian (note the logarithmic scale in Fig. 1b). Consequently, the median, 0.195 d^{-1} , is more representative. For reference, the geometric mean is 0.162 d^{-1} . The combined data set is dominated by the 274 observations coming from the POMME programme in the northeast Atlantic²⁶. Excluding these data, the

¹National Oceanography Centre, Southampton, European Way, Southampton SO14 3ZH, UK. ²Laboratoire d’Océanographie et de Biogéochimie, Centre d’Océanologie de Marseille, 163 avenue de Luminy Case 901, F-13288 Marseille, France. ³Laboratorio de Procesos Oceanográficos y Clima, Departamento de Oceanografía y Centro de Investigación Oceanográfica en el Pacífico Sur Oriental, Universidad de Concepción, Casilla 160-C, Concepción, Chile. ⁴Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, UK.

median is 0.066 d^{-1} (the POMME data have a median of 0.302 d^{-1}). Although such a bias might suggest that the POMME data should be excluded, this is unwarranted because they comprise the largest pool of observations. The median of the full set is therefore held to be the best representative value.

There is no indication of significant relationships between λ_{nitrif} and either the time of year or the latitude (see Supplementary Information). This does not imply that such relationships do not exist; rather, it indicates that they might be obscured by variability arising from current observational constraints and the paucity of observations.

The magnitude and vertical distribution of λ_{nitrif} collated in Fig. 1b strongly indicate that nitrification is a very important factor in the euphotic nitrogen cycle. Complete turnover of the ammonium pool in less than 24 h may be possible, even in surface waters. However, determining the large-scale significance of these measurements is less straightforward, because there are still relatively few of them, and they are limited in space and time.

To attempt to quantify the consequences of these observations for ‘new’ production, a model incorporating nitrification was employed to examine the process at the global scale. This model uses an ocean general circulation model in conjunction with a nutrient–phytoplankton–zooplankton–detritus biogeochemical scheme. This treats nitrate and ammonium separately, and additionally models ‘regenerated’ nitrate (arising from the nitrification of ammonium) as a separate tracer to distinguish it from ‘new’ nitrate (coming from depth, as defined below). Uptake of these three nutrients is also separately tracked. Nitrification is modelled simply as the ‘decay’ of

ammonium to ‘regenerated’ nitrate at a constant specific rate—the simplest assumption, given the current uncertainties related to the process. On the basis of the data in Fig. 1 and the argument given above, we take λ_{nitrif} to be 0.2 d^{-1} . ‘Regenerated’ nitrate is converted to ‘new’ nitrate only beneath the ‘permanent thermocline’ (as defined in Supplementary Information), where it happens instantaneously. The nitrogen-cycle processes of denitrification, riverine inputs and nitrogen fixation are not included in this model. The model was run for a period of 20 years (1985–2004); the final decade was used to analyse model behaviour. A full description of the model is given in Supplementary Information.

Figure 2a shows annual average total primary production. Although the modelled field has some deficiencies (for example, excessive production in equatorial upwelling regions), its distribution of production is otherwise comparable to observations, and its global total, 51 Gt C yr^{-1} , is close to estimates²⁷. Figure 2b shows the classic *f*-ratio, calculated by separating production fuelled by ammonium from that fuelled by nitrate (the sum of both ‘new’ and ‘regenerated’ nitrate here). About half of phytoplankton production is fuelled by nitrate, with the *f*-ratio increasing towards higher latitudes, where a shorter growing season limits nitrate depletion. Equatorial upwelling regimes, where nitrate is supplied continuously from depth, break this general geographical pattern by having higher *f*-ratios than neighbouring low latitudes. The *f*-ratio reaches its lowest values in the oligotrophic gyre regions, where stratification keeps ‘new’ production low and recycling high.

By differentiating ‘regenerated’ nitrate from ‘new’ nitrate, it is possible to recalculate the *f*-ratio by using only ‘new’ nitrate. As Fig. 2c shows, on the global scale less than half of the production previously identified as ‘new’ is fuelled by nitrate from below the

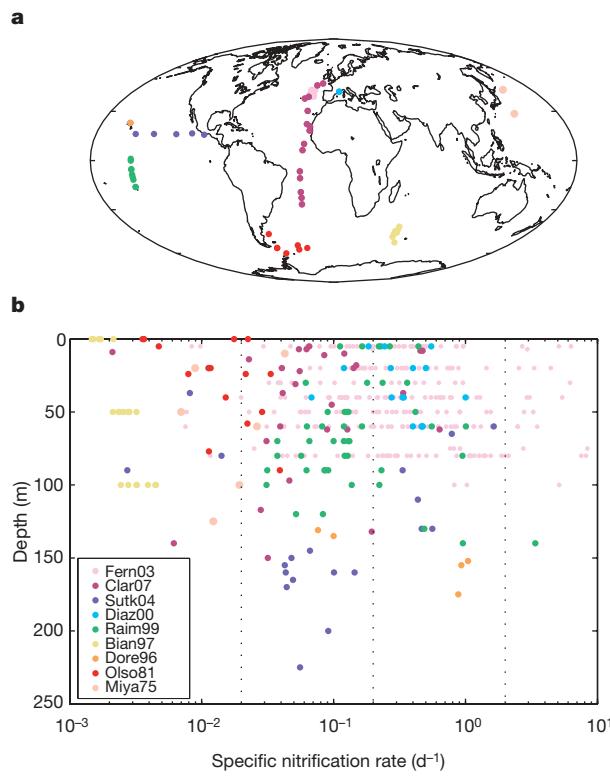


Figure 1 | Global distribution of nitrification observations and the relationship with depth. Collated measurements of nitrification from a series of observational studies, plotted to show study sites (a) and the relation between specific nitrification rate (d^{-1}) and depth (b). The dotted vertical lines in b indicate modelled nitrification rates in default (centre) and sensitivity analysis (left and right) simulations. In chronological order, the cited studies are Miya75 (ref. 4), Ols081 (ref. 5), Dore96 (ref. 6), Bian97 (ref. 7), Raim99 (ref. 8), Diaz00 (ref. 9), Fern03 (ref. 10), Sutk04 (ref. 11) and Clar07 (ref. 12); however, because it dominates the data set, Fern03 is plotted out of sequence for clarity.

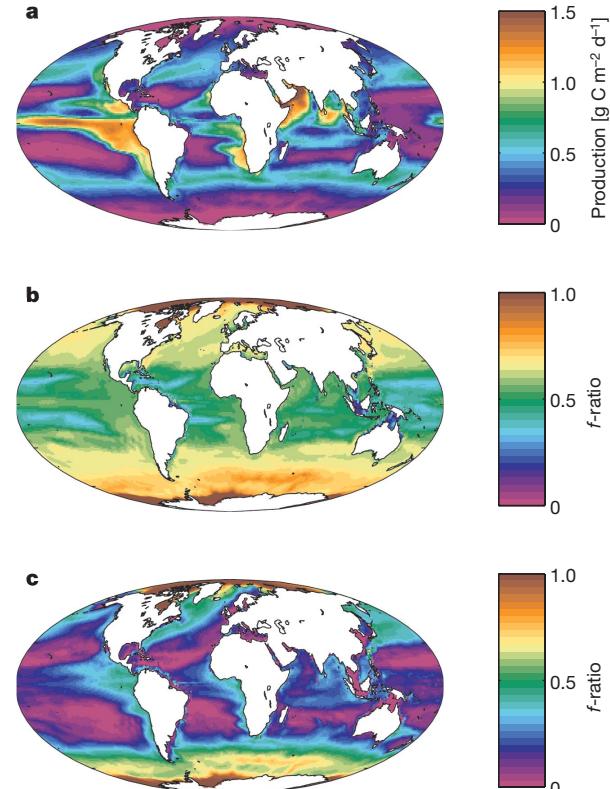


Figure 2 | Modelled primary production and *f*-ratio. a, Modelled primary production, averaged for the simulation period 1995–2004 (global average $51.24 \text{ Gt C yr}^{-1}$). b, c, *f*-ratio averaged for the same period. b, The classic *f*-ratio (global average 0.577), the fraction of production fuelled by both ‘new’ and ‘regenerated’ nitrate. c, The ‘new’-only *f*-ratio (global average 0.260), the fraction fuelled solely by ‘new’ nitrate (like ammonium, ‘regenerated’ nitrate is considered ‘regenerated’ in this plot).

Table 1 | Summary of default and sensitivity simulations

λ_{nitrif} (d^{-1})	Primary production (Gt C yr^{-1})			f -ratio	$N_{\text{reg}} (\%)$		
	Total	N_{new}	N_{reg}	$N_{\text{new}} + N_{\text{reg}}$			
0.02	50.88	11.73	6.28	0.354	0.231	0.327	18.69
0.20	51.26	13.32	16.26	0.577	0.260	0.461	27.07
2.00	51.56	14.16	30.44	0.865	0.275	0.625	32.81

Integrated primary production, f -ratio and 'regenerated' nitrate fraction in the default simulation and sensitivity analyses examining the effect of the magnitude of nitrification rate, λ_{nitrif} , are given. Primary production is shown for all nutrient sources, for 'new' nitrate only and for 'regenerated' nitrate only. The f -ratios shown are for classic production, for 'new'-only-fuelled production and for 'corrected' ($N_{\text{new}} + N_{\text{nit}}$ – nitrification)-fuelled production. The final column shows the average fraction of total surface nitrate that is 'regenerated' (not 'new') in the simulations.

permanent thermocline. Regionally, excluding 'regenerated' nitrate causes even larger changes in the f -ratio, especially in subtropical gyres. In these regions, a formerly even split between 'new' and 'regenerated' production is instead revealed to be overwhelmingly dominated by recycled nitrogen. The regions where the f -ratio changes least, namely the high latitudes and upwelling regions, are those identified earlier as experiencing short growing seasons or continuous upwelling of deep (and hence 'new') nitrate.

Given the strong variability in the data it is important to determine the sensitivity of these results to our chosen nitrification rate. Table 1 summarizes estimates of primary production and f -ratios from two additional simulations (see Supplementary Information) in which the rate of nitrification is decreased and increased by an order of magnitude, broadly encompassing the range in Fig. 1b. In both simulations, whereas the classic f -ratios shift markedly from that of the default simulation, the 'new'-only f -ratios shift only slightly. However, the discrepancy between classic and 'new'-only f -ratios remains, even when the nitrification rate is radically decreased.

These model results suggest that nitrification substantially distorts estimates of 'new' production (and therefore export production) when based on bulk nitrate uptake. Although these distortions are greatest in the relatively unproductive oligotrophic regions (where 'regenerated' nitrate can form more than 70% of the total euphotic zone nitrate pool), the wider results suggest that estimates of 'new' production based solely on the relative uptake rates of nitrate and ammonium may significantly overestimate 'new' production at the global scale. Furthermore, no simple local correction to account for nitrification exists. Table 1 shows that subtracting measured nitrification from nitrate uptake²⁶ provides an insufficient correction (see Supplementary Information).

A caveat to this work is that, although nitrification rates have been observationally quantified then applied linearly in a simple model, the factors that influence the process itself are still not fully understood. For instance, nitrification occurs in two stages mediated by two different groups of bacteria, opening the possibility of niche separation. Furthermore, other groups, such as the Crenarchaeota, have only recently been identified as nitrifiers²⁸. As well as ecophysiological factors, community structure may have a role because phytoplankton can compete with nitrifiers for ammonium in well-lit ocean layers²⁹. However, strong relationships between nitrification and environmental or ecological factors have yet to be established²⁹.

Given the strong interest in the biological pump and its role in the ocean's future ability to sequester anthropogenic CO₂, these findings undermine one of the most popular methods for estimating its strength. Although the difficulty posed by nitrification in establishing the provenance of nitrate was highlighted when the f -ratio was originally defined², measurement of the latter subsequently entered the standard suite of oceanographic techniques. There are still many ecological systems in which separating nitrate and ammonium use is important³⁰, but the blanket use of the f -ratio to diagnose export production no longer seems either justified or practical.

METHODS SUMMARY

The field studies synthesized in this work^{4–12} measured nitrification by using a range of different techniques and experimental designs. Most studies used the

natural¹¹ or 'spiked'^{4,5,8–10,12} abundance of the ¹⁵N isotope, and tracked this from the ammonium to the nitrite and/or nitrate pools. Quantification was achieved either by isotope ratio mass spectrometry^{4,5,8–11} or gas chromatography–mass spectrometry¹². One study⁶ estimated nitrification by quantifying the carbon assimilated by nitrifying bacteria (using ¹⁴C), and then calibrating the C/N ratio of this assimilation by means of a simple chemical assay. The remaining study⁷ estimated nitrification by running parallel incubations, one of which was poisoned with allylthiourea to inhibit ammonium oxidation.

All except one of the data sets measured the rate of ammonium oxidation and ammonium concentration simultaneously^{4,5,7–12}, so we defined the specific nitrification rate, λ_{nitrif} , as the ammonium oxidation rate divided by the corresponding ammonium concentration. This assumed that the specific oxidation rate of nitrite was at least equal to that of ammonium. This was true for 85% of the observations in which both oxidation rates were measured. For the eighth data set⁶ we estimated the specific rate by dividing nitrite oxidation rate by nitrite concentration.

The physical model used was OCCAM¹³, a medium-resolution primitive-equation ocean general circulation model (vertical resolution 66 levels; horizontal resolution 1°). OCCAM includes an elastic–viscous–plastic sea-ice scheme, a K -profile parameterization mixed layer, and Gent–McWilliams eddy parameterization. Surface fluxes are calculated with empirical formulae and NCEP-derived atmospheric boundary-layer quantities (1985–2004). The ecosystem model used is a variant of an existing nutrient–phytoplankton–zooplankton–detritus model¹⁴ in which nitrogenous nutrient has been separated into 'new' nitrate, ammonium and 'regenerated' nitrate. Nitrification occurs throughout the water column, but 'regenerated' nitrate only becomes 'new' nitrate below a regionally variable horizon defined by the maximum mixed-layer depth during the simulation.

Received 16 February; accepted 27 April 2007.

- Raven, J. A. & Falkowski, P. G. Oceanic sinks for atmospheric CO₂. *Plant Cell Environ.* 22, 741–755 (1999).
- Eppley, R. W. & Peterson, B. J. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* 282, 677–680 (1979).
- Dugdale, R. C. & Goering, J. J. Uptake of new and regenerated forms of nitrogen in primary production. *Limnol. Oceanogr.* 12, 196–206 (1967).
- Miyazaki, T., Wada, E. & Hattori, A. Nitrite production from ammonia and nitrate in the euphotic layer of the western North Pacific Ocean. *Mar. Sci. Commun.* 1, 381–394 (1975).
- Olson, R. J. ¹⁵N tracer studies of the primary nitrite maximum. *J. Mar. Res.* 39, 203–226 (1981).
- Dore, J. E. & Karl, D. M. Nitrification in the euphotic zone as a source for nitrite, nitrate, and nitrous oxide at Station ALOHA. *Limnol. Oceanogr.* 41, 1619–1628 (1996).
- Bianchi, M., Feliatra, F., Tréguer, P., Vincedieu, M.-A. & Morvan, J. Nitrification rates, ammonium and nitrate distribution in upper layers of the water column and in sediments of the Indian sector of the Southern Ocean. *Deep-sea Res. II* 44, 1017–1032 (1997).
- Rainbault, P. et al. Carbon and nitrogen uptake and export in the equatorial Pacific at 150°W: Evidence of an efficient regenerated production cycle. *J. Geophys. Res.* 104, 3341–3356 (1999).
- Diaz, F. & Rainbault, P. Nitrogen regeneration and dissolved organic nitrogen release during spring in a NW Mediterranean coastal zone (Gulf of Lions): implications for the estimation of new production. *Mar. Ecol. Prog. Ser.* 197, 51–65 (2000).
- Fernández, C. *Cycle de l'Azote et Production Primaire dans l'Atlantique Nord-Est: Suivi Saisonnier et Influence de la Meso Échelle*. PhD thesis, Univ. de la Méditerranée, Marseille (2003).
- Sutka, R. L., Ostrom, N. E., Ostrom, P. H. & Phanikumar, M. S. Stable nitrogen isotope dynamics of dissolved nitrate in a transect from the North Pacific Subtropical Gyre to the Eastern Tropical North Pacific. *Geochim. Cosmochim. Acta* 68, 517–527 (2004).
- Clark, D. R., Rees, A. P. & Joint, I. A method for the determination of nitrification rates in oligotrophic marine seawater by gas chromatography/mass spectrometry. *Mar. Chem.* 103, 84–96 (2007).
- Marsh, R., de Cuevas, B. A., Coward, A. C. & Bryden, H. L. and Alvarez, M. Thermohaline circulation at three key sections of the North Atlantic over 1985–2002. *Geophys. Res. Lett.* 32, L10604, doi:10.1029/2004GL022281 (2005).
- Oschlies, A. NAO-induced long-term changes in nutrient supply to the surface waters of the North Atlantic. *Geophys. Res. Lett.* 28, 1751–1754 (2001).
- Takahashi, T. et al. Global sea-air CO₂ flux based on climatological surface ocean pCO₂, and seasonal biological and temperature effects. *Deep-sea Res. II* 49, 1601–1622 (2002).
- Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S. & Wakeham, S. G. A new, mechanistic model for organic carbon fluxes in the ocean: based on the quantitative association of POC with ballast minerals. *Deep-sea Res. II* 49, 219–236 (2002).

17. Key, R. M. et al. A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochem. Cycles* **18**, GB4031, doi:10.1029/2004GB002247 (2004).
18. Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).
19. Orr, J. C. et al. Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* **437**, 681–686 (2005).
20. Olson, R. J. Differential photoinhibition of marine nitrifying bacteria - a possible mechanism for the formation of the primary nitrite maximum. *J. Mar. Res.* **39**, 227–238 (1981).
21. Guerrero, M. A. & Jones, R. D. Photoinhibition of marine nitrifying bacteria 1. Wavelength-dependent response. *Mar. Ecol. Prog. Ser.* **141**, 183–192 (1996).
22. Ward, B. B., Kilpatrick, K. A., Renger, E. H. & Eppley, R. W. Biological nitrogen cycling in the nitracline. *Limnol. Oceanogr.* **34**, 493–513 (1989).
23. Slawyk, G. & Raimbault, P. Simple procedure for simultaneous recovery of dissolved inorganic and organic nitrogen in ^{15}N -tracer experiments and improving the isotopic mass-balance. *Mar. Ecol. Prog. Ser.* **124**, 289–299 (1995).
24. Lipshultz, F. A time-series assessment of the nitrogen cycle at BATS. *Deep-sea Res. II* **48**, 1897–1924 (2001).
25. Martin, A. P. & Pondaven, P. New primary production and nitrification in the western subtropical North Atlantic: a modelling study. *Glob. Biogeochem. Cycles* **20**, 10.1029/2005GB002608 (2006).
26. Fernández, C., Raimbault, P., Garcia, N., Rimmelin, P. & Caniaux, G. An estimation of annual new production and carbon fluxes in the northeast Atlantic Ocean during 2001. *J. Geophys. Res. Oceans* **110**, C07S13 (2005).
27. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
28. Wuchter, C. et al. Archaeal nitrification in the ocean. *Proc. Natl Acad. Sci. USA* **103**, 12317–12322 (2006).
29. Ward, B. Temporal variability in nitrification rates and related biogeochemical factors in Monterey Bay, California, USA. *Mar. Ecol. Prog. Ser.* **292**, 97–109 (2005).
30. Allen, A. E. et al. Importance of heterotrophic bacterial assimilation of ammonium and nitrate in the Barents Sea during summer. *J. Mar. Syst.* **38**, 93–108 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Sinha, B. de Cuevas, G. Nurser and A. Coward for technical assistance with OCCAM; A. Rees for discussions of nitrification; A. Oschlies for comments on the manuscript; and P. Raimbault and R. Sutka for providing nitrification data. PATOM/PROOF (France) provided funding for the POMME experimental programme, the largest source of nitrification observations so far. A.Y. is funded by a UK National Environment Research Council (NERC) standard grant; A.P.M. is funded by a NERC Advanced Research fellowship; C.F. is funded by Fundación Andes, Chile; and D.R.C. was supported by a NERC grant and through the Atlantic Meridional Transect consortium.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.Y. (axy@noc.soton.ac.uk).

LETTERS

Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary

J. R. Wible¹, G. W. Rougier², M. J. Novacek³ & R. J. Asher⁴

Estimates of the time of origin for placental mammals from DNA studies span nearly the duration of the Cretaceous period (145 to 65 million years ago), with a maximum of 129 million years ago¹ and a minimum of 78 million years ago². Palaeontologists too are divided on the timing. Some^{3–5} support a deep Cretaceous origin by allying certain middle Cretaceous fossils (97–90 million years old) from Uzbekistan with modern placental lineages, whereas others^{6,7} support the origin of crown group Placentalia near the close of the Cretaceous. This controversy has yet to be addressed by a comprehensive phylogenetic analysis that includes all well-known Cretaceous fossils and a wide sample of morphology among Tertiary and recent placentals⁶. Here we report the discovery of a new well-preserved mammal from the Late Cretaceous of Mongolia and a broad-scale phylogenetic analysis. Our results exclude Cretaceous fossils from Placentalia, place the origin of Placentalia near the Cretaceous/Tertiary (K/T) boundary in Laurasia rather than much earlier within the Cretaceous in the Southern Hemisphere^{8,9}, and place afrotherians and xenarthrans in a nested rather than a basal position^{8,9} within Placentalia.

Placentals represent most living mammals (1,135 out of 1,229 genera) and are found on all continents and in all oceans¹⁰. Placentals and their extinct stem lineage constitute the Eutheria. More than 4,000 extinct eutherian genera have been named that represent the 65 million years of the Cenozoic, and the majority of these have been assigned to modern placental lineages¹¹. In contrast, only about 40 eutherian genera are known from the 80 million years of the Cretaceous^{4,6,11}. The relationships of these Cretaceous taxa to modern placentals are highly contentious. At one extreme, three-quarters of Cretaceous eutherians are assigned to the placental crown group^{4,11}; at the other, all Cretaceous eutherians fall outside Placentalia⁶.

Mammalia Linnaeus, 1758

Theria Parker and Haswell, 1897

Eutheria Gill, 1872

Cimolestidae Marsh, 1889

Maelestes gobiensis gen. et sp. nov.

Etymology. *Mae* is the acronym for Mongolian Academy of Sciences–American Museum of Natural History Expeditions; *lestes* (Greek), robber, often used for insectivore-like mammals; *gobiensis*, occurring in the Gobi Desert.

Holotype. PSS-MAE 607 (Figs 1 and 2; Palaeontological and Stratigraphy Section, Geological Institute, Mongolian Academy of Sciences, Ulaan Baatar): an incomplete skull, left mandible, atlas, axis, twelve thoracic vertebrae, eight partial ribs, incomplete scapula, clavicle, humerus, proximal radius and ulna, and incomplete astragalus (see Supplementary Figs 1–5).

Age and locality. Late Cretaceous Djadokhta Formation, Ukhaa Tolgod (between Camel Humps and Sugar Mountain), Mongolia. Recent age estimates of Djadokhta Formation are 75–71 million years¹².

Diagnosis. Upper dentition: I², C¹, P⁵, M³. Lower dentition: I₃, C₁, P₅, M₃ (Figs 1 and 2). Differs from other Mongolian Djadokhta Formation eutherians (*asioryctitheres* *Kennalestes*, *Asioryctes*¹³ and *Ukhaatherium*¹⁴, and zalmabdalestsids *Zalambdalestes* and *Barunlestes*¹⁵) in having five upper and lower premolars, three subequal procumbent lower incisors, palatal vacuity between maxilla and palatine, postglenoid foramen behind postglenoid process, transmontriorial internal carotid artery, and small pectoral canal. Resembles *Kennalestes* and *Asioryctes*¹³ in having hypoglossal foramen housed in an opening larger than jugular foramen and petrosal roof for external acoustic meatus. Differs from Central Asian Late Cretaceous *asioryctitheres* *Bulaklestes*, *Daulestes* and *Uchukudukodon*¹⁶ in having five upper and lower premolars, single-rooted lower canine, penultimate upper premolar with three roots, upper molars much wider than long with narrower stylar shelves, and lower molars with protoconid subequal to metaconid. Differs from Central Asian Late Cretaceous ‘zhelestids’ (*Sheikhdzheilia*, *Eozhelestes*, *Aspanlestes*, *Zhelestes* and *Parazhelestes*)^{3,17,18} and North American and Central Asian Late Cretaceous *Paranyctoides*^{3,18} in having upper molars with weak conules, metacone much smaller than paracone, metacone and paracone with adjoined base, and lower molars with narrower talonids and no labial postcingulid. Resembles North American Late Cretaceous cimolestids⁴ *Cimolestes* and *Batodon*^{19,20} in having lower canine and first lower premolar single-rooted. Resembles *Cimolestes* in having subequal procumbent lower incisors (two preserved in *Cimolestes propalaeoryctes*)¹⁹. Resembles *Batodon*^{18,19} in having upper molars with narrow stylar shelves and pre- and post-cingula, and lower molars with transverse protocristid and entoconid approximating hypoconulid. Differs from *Cimolestes* and *Batodon*^{19,20} in having five upper and lower premolars, upper molars with weak conules, and lower molars with more compressed trigonids and protoconid subequal to metaconid.

Our phylogenetic analysis (see below) allies *Maelestes* with two slightly younger western North American taxa, the cimolestids⁴ *Cimolestes* and *Batodon* (Fig. 3); these are known primarily by incomplete dentitions and jaws^{4,19,20}, and have been linked with placental carnivorans⁴. *Maelestes* is the sister of *Batodon*, the smallest Cretaceous eutherian, which has molars roughly 60% the size of *Maelestes*. The dentition of *Maelestes* shows an odd mix of resemblances to other Late Cretaceous Asian taxa, with premolars like the zhelestid *Zhelestes*¹⁸, upper molars like the asioryctithere *Kennalestes*¹³, and lower molars like the zalmabdalestid *Zalambdalestes*¹⁵. *Maelestes* is the first Cretaceous eutherian with a marsupial-like palatal vacuity (Fig. 1)—a rare feature, even among extant placentals (for example, some

¹Section of Mammals, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15206, USA. ²Department of Anatomical Sciences and Neurobiology, School of Medicine, University of Louisville, Louisville, Kentucky 40292, USA. ³Division of Paleontology, American Museum of Natural History, New York, New York 10024, USA. ⁴Museum of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK.

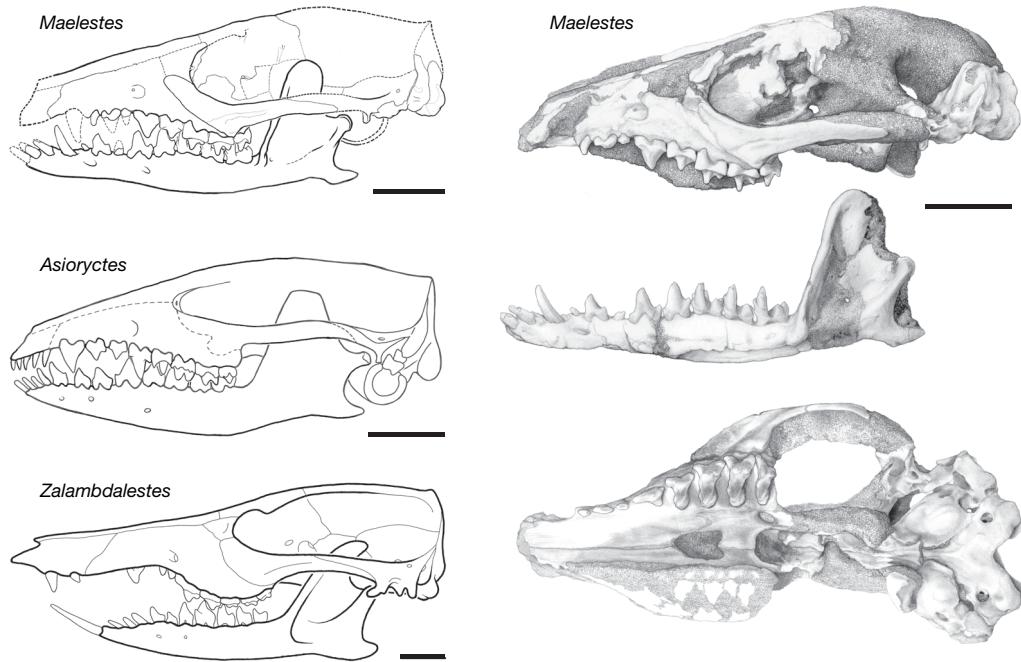


Figure 1 | *Maelestes gobiensis* gen. et sp. nov. (holotype, PSS-MAE 607) skull and mandible in comparison to other Djadokhta eutherians. Left panel: reconstructions of *Maelestes*, *Asioryctes* and *Zalambdalestes*, the last two of which are redrawn from ref. 15. Right panel: incomplete *Maelestes* skull in left lateral (top) and ventral (bottom) views, and left *Maelestes*

hedgehogs and elephant shrews). The preserved postcranial elements of *Maelestes* are similar to those of the asioryctithere *Ukhaatherium* from Ukhaa Tolgod, which resemble, but are more generalized than, those of placental insectivores²¹.

We reconstructed the phylogenetic relationships of *Maelestes* by parsimony analysis of 408 morphological characters (127 dental, 212 craniomandibular and 69 postcranial) across 69 taxa (see

mandible in lateral view (middle). The large opening in the palate between the palatines and maxillae is a palatal vacuity. *Maelestes* probably had some upper incisors, but only a small non-tooth-bearing fragment of the premaxilla is preserved. Scale bars, 5 mm.

Supplementary Information), including 4 stem therians, 3 metatherians, 31 Cretaceous eutherians (all but the most incomplete and poorly preserved taxa), 20 extinct Tertiary placentals and 11 extant placentals. The Tertiary and extant taxa were chosen to sample the 4 major placental lineages recovered by some recent DNA studies^{8,9}: 5 afrotherians, 3 xenarthrans, 10 euarchontaglians and 13 laurasiatherians. We did not include any of the Jurassic and Cretaceous Gondwanan mammals (*Ambondro*, *Asfaltomylos*, *Ausktribosphenos* and *Bishops*), which are regarded by some as eutherians²², because most recent analyses place these taxa in a Southern Hemisphere clade, Australosphenida, that is more distantly related to placentals than the stem therians and metatherians used here as outgroups^{4,23}.

Our strict consensus tree (Fig. 3) recognizes at least nine lineages of Mesozoic eutherians and does not support the inclusion of any Cretaceous eutherians within a placental lineage. Although branch support for many basal nodes is weak, we examined various competing hypotheses of association^{3–6,11} using a Wilcoxon rank sum (also known as Templeton) test, and found all (except *Purgatorius* with Primates, and Palaeocene and Eocene ‘condylarths’ with Cetartiodactyla) to be significantly rejected using our morphological data set (see Supplementary Information). Of the five basal-most eutherians in our tree, all are from Asia except *Montanalestes*, which is from western North America; this supports an Asian origin of Eutheria and its sister group Metatheria (marsupials and their stem lineage), because the basal-most members of both clades are Asian⁴.

Most Late Cretaceous eutherians fall into three morphologically distinct clades. The basal-most of these, the Zhelestidae (Fig 3), is the most widespread temporally and geographically, occurring in Uzbekistan, western North America and Spain. Zhelestids are known mainly from their incomplete dentitions^{3,4,17,18}. They have robust upper molar protocones, shifted labially in some forms, which are among the features used to support the view that zhelestids are a paraphyletic stem lineage to ‘condylarths’ (basal ungulates)^{3–5,18}. In contrast, our analysis (Fig. 3) shows that upper molar resemblances were acquired convergently in zhelestids and ‘condylarths’, and that a minimum of 25 additional steps are required to produce a zhelestid–‘condylarths’

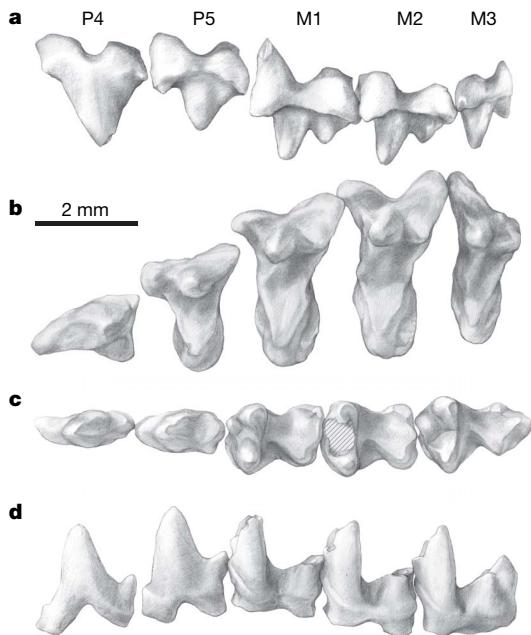


Figure 2 | *Maelestes gobiensis* gen. et sp. nov. (holotype, PSS-MAE 607) left upper and lower penultimate and ultimate premolars (P4, P5) and molars (M1, M2, M3). Uppers are shown in labial (a) and occlusal (b) views; lowers are shown in occlusal (c) and labial (d) views.

clade. Furthermore, our results (Fig. 3) highlight the need for representative taxonomic as well as morphological sampling in phylogenetic analyses, because the two ‘condylarths’ from the early Palaeocene (*Protungulatum* and *Oxyprimus*), aligned with zhelestids elsewhere^{3,4,18}, do not appear with late Palaeocene–early Eocene ‘condylarths’ (*Hyopsodus*, *Meniscotherium* and *Phenacodus*) or even within the placental crown group. Our results support *Hyopsodus*, *Meniscotherium* and *Phenacodus* as the oldest ungulate-like clade of crown placentals, which form the sister clade to the early Eocene dichobunid cetartiodactylan *Gujaratia pakistanensis*. Other morphological analyses have placed North American ‘condylarths’ with Afroteria^{24,25}, which adds an extra 30 steps to our optimal trees.

The other two Late Cretaceous clades include some forms represented by fairly complete skulls and skeletons. Zalambdalestidae is the more proximate of the two to Placentalia (Fig. 3). Zalambdalestids are endemic to Asia; they are dentally specialized with enlarged evergrowing anteriormost lower incisors (*Zalambdalestes* in Fig. 1) that have enamel restricted anteriorly, as occurs in Glires (rodents and lagomorphs)^{3,15} as well as in various other mammals¹⁵. This dental specialization has been used elsewhere³ to support a zalambdalestid–Glires clade. In contrast, our analysis (Fig. 3) indicates that the specialized lower incisors of Zalambdalestidae and Glires were acquired convergently (confirming some earlier results^{15,26}). An extra 35 steps from the most parsimonious trees are needed to place zalambdalestids with Glires. The third Late Cretaceous clade is dentally more generalized than the other two. Included are Asioryctitheria^{14,16} (*Asioryctes* in Fig. 1), endemic to Asia, together with the cimolestid lineage that contains *Maelestes*, *Batodon* and *Cimolestes* (Fig. 3).

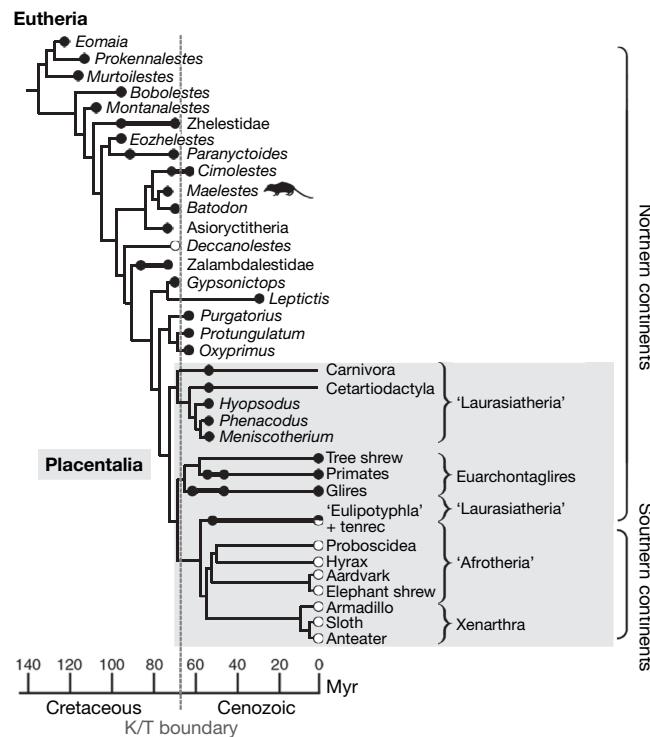


Figure 3 | Relationships of *Maelestes* to other eutherians. Simplified eutherian part of the strict consensus of three most parsimonious trees (2,296 steps; see Supplementary Information). Broken line, K/T boundary at 65 million years. Filled circle taxa, Laurasian (northern continents); open circle taxa, Gondwanan (southern continents). Circles and circles with thicker black lines indicate temporal occurrence of studied specimens. The grey box delimits Placentalia, the oldest member of which is the early Palaeocene *Mimotona*. Placentalia originates at or near the K/T boundary in Laurasia. *Deccanolestes* from the latest Cretaceous of India is nested among Asian clades and is more parsimoniously interpreted as an independent southern migration. The animal silhouette spotlights *Maelestes*.

Three models of origin and diversification of the modern placental orders have been characterized²⁷: ‘explosive’ (interordinal and ordinal origins near the K/T boundary); ‘long fuse’ (interordinal divergence deep in the Cretaceous with intraordinal diversification near the K/T boundary); and ‘short fuse’ (interordinal, ordinal and intraordinal diversifications deep in the Cretaceous, but without a fossil record). Some recent molecular studies^{2,8,9,28} are consistent with the short- or long-fuse models (deep Cretaceous origin), and a recent supertree approach²⁹ is consistent with the short-fuse model, although it has been argued that this model is statistically unlikely^{7,27}. Some palaeontologists^{6,7} prefer the explosive model (K/T boundary origin and diversification), despite the paucity of supporting phylogenetic analyses. Our analysis, which to date provides the best sample of relevant taxa and morphological characters, supports the explosive model. The immediate outgroup to Placentalia is a clade of *Protungulatum*, *Oxyprimus* and *Purgatorius*, best known from the Bug Creek Anthills of Montana, originally reported as latest Cretaceous but now accepted as basal Palaeocene⁴. All known fossil placentals are younger than this (Fig. 3), although some extension of ghost lineages into the Cretaceous is likely.

Within Placentalia, our results (Fig. 3) identify Euarchontaglires and Xenarthra—two of the four major placental lineages identified in some recent DNA studies^{8,9}. Our morphological dataset does not fully support the other two lineages, Afroteria and Laurasiatheria, as monophyletic, with the afroterian tenrec *Potamogale* nested within laurasiatherian Eulipotyphla and eulipotyphlans more closely related to xenarthrans and the remaining afroterians than to other laurasiatherians (carnivorans and cetartiodactylans). Afroteria is a novel molecular clade^{8,9} that unites the modern placental clades Hyracoidea (hyraxes), Proboscidea (elephants), Sirenia (manatees), Tubulidentata (aardvarks), Macroscelidea (elephant shrews), Tenrecidae (tenrecs) and Chryschloridae (golden moles)—groups usually aligned with other ungulate- and insectivore-like lineages in morphological taxonomies¹¹. Our analysis, however, captures a core Afroteria (hyrax, the Eocene proboscidean *Moeritherium*, aardvark and elephant shrew).

The highly nested position for Afroteria and Xenarthra (the South American clade of armadillos, sloths and anteaters) in our tree (Fig. 3) is strikingly different from most recent molecular results^{8,9} (but resembles mitogenomic analyses²⁸ as well as from combined molecular and morphological results²⁴), which identify these groups as the basalmost placental clades, supporting a Southern Hemisphere origin for Placentalia. The nested position for Xenarthra is also a departure from previous classifications that place this clade at the base of Placentalia¹¹. Given that the basal placental clades and immediate outgroups in our tree are from the Northern Hemisphere (Fig. 3), our analysis supports a laurasian origin for Placentalia, with subsequent appearance of afroterians as recovered here in Africa and xenarthrans in South America. A laurasian origin has been argued as most parsimonious even if afroterians and xenarthrans are at the base of Placentalia⁷. The only Cretaceous Gondwanan form in our tree—*Deccanolestes* from the latest Cretaceous of India—is nested among Asian clades and, therefore, is most parsimoniously an immigrant from the north, as proposed elsewhere³⁰.

Elucidating the origin and diversification of placentals is facilitated by palaeontological and neontological approaches. We see the progressive convergence of tree topologies resulting from recent molecular, morphological and combined data sets as highly encouraging. However, the temporal framework for the origination of Placentalia ranges, potentially, from the Early Cretaceous to the early Palaeocene. Some recent molecular clock studies^{8,9,28} and a recent supertree approach²⁹ support origin dates for Placentalia close to 100 million years. But other clock studies vary widely for this node, with dates that exceed 140 million years¹ or that approach palaeontological estimates of less than 80 million years². Other analyses^{7,26}, including our own, support an origin in the vicinity of the K/T boundary. These competing hypotheses imply substantially different palaeogeographical

scenarios that have crucial impacts on the areas of origin, dispersion and diversification of some of the major clades of living mammals.

Received 21 February; accepted 17 April 2007.

1. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
2. Douzery, E. J. P., Delsuc, F., Stanhope, M. J. & Huchon, D. Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals and incompatibility among fossil calibrations. *J. Mol. Evol.* **57**, S201–S213 (2003).
3. Archibald, J. D., Averianov, A. O. & Ekdale, E. G. Oldest relative to Glires and the Late Cretaceous roots of Placentalia. *Nature* **414**, 62–65 (2001).
4. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs: Origins, Evolution, and Structure* (Columbia Univ. Press, New York, 2004).
5. Benton, M. J. & Donoghue, P. C. J. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26–53 (2007).
6. Wible, J. R., Rougier, G. W. & Novacek, M. J. in *The Rise of Placental Mammals: Origins and Relationships of the Major Extant Clades* (eds Rose, K. D. & Archibald, J. D) 15–36 (Johns Hopkins Univ. Press, Baltimore, 2005).
7. Hunter, J. P. & Janis, C. M. Spiny Norman in the Garden of Eden? Dispersal and early biogeography of Placentalia. *J. Mammal. Evol.* **13**, 89–123 (2006).
8. Murphy, W. J. et al. Resolution of early placental mammal radiation using Bayesian phylogenetics. *Science* **294**, 2348–2351 (2001).
9. Springer, M. S., Murphy, W. J., Eizirik, E. & O'Brien, S. J. in *The Rise of Placental Mammals: Origins and Relationships of the Major Extant Clades* (eds Rose, K. D. & Archibald, J. D) 37–49 (Johns Hopkins Univ. Press, Baltimore, 2005).
10. Wilson, D. E. & Reeder, D. M. (eds) *Mammal Species of the World: A Taxonomic and Geographic Reference* (Johns Hopkins Univ. Press, Baltimore, 2005).
11. McKenna, M. C. & Bell, S. K. *Classification of Mammals Above the Species Level* (Columbia Univ. Press, New York, 1997).
12. Dashzeveg, D. et al. New stratigraphic subdivision, depositional environment, and age estimate for the Upper Cretaceous Djadokhta Formation, southern Ulan Nur Basin, Mongolia. *Am. Mus. Novitates* **3498**, 1–31 (2005).
13. Kielan-Jaworowska, Z. Evolution of the therian mammals in the Late Cretaceous of Asia. Part IV. Skull structure in *Kennalestes* and *Asioryctes*. *Palaeontol. Pol.* **42**, 25–78 (1981).
14. Novacek, M. J. et al. Epipubic bones in eutherian mammals from the Late Cretaceous of Mongolia. *Nature* **389**, 483–486 (1997).
15. Wible, J. R., Novacek, M. J. & Rougier, G. W. New data on the skull and dentition in the Mongolian Late Cretaceous eutherian mammal *Zalambdalestes*. *Bull. Am. Mus. Nat. Hist.* **281**, 1–144 (2004).
16. Archibald, J. D. & Averianov, A. O. Late Cretaceous asioryctitherian eutherian mammals from Uzbekistan and phylogenetic analysis of Asioryctitheria. *Acta Palaeontol. Pol.* **51**, 351–376 (2006).
17. Averianov, A. & Archibald, J. D. Mammals from the mid-Cretaceous Khokzhakul Formation, Kyzylkum Desert, Uzbekistan. *Cretac. Res.* **26**, 593–608 (2005).
18. Nessov, L. A., Archibald, J. D. & Kielan-Jaworowska, Z. Ungulate-like mammals from the Late Cretaceous of Uzbekistan and a phylogenetic analysis of Ungulatomorpha. *Bull. Carnegie Mus. Nat. Hist.* **34**, 40–88 (1998).
19. Lillegraven, J. A. Latest Cretaceous mammals from the upper part of the Edmonton Formation of Alberta, Canada, and review of marsupial-placental dichotomy in mammalian evolution. *Univ. Kansas Paleontol. Contrib.* **50**, 1–122 (1969).
20. Clemens, W. A. Jr. Fossil mammals of the type Lance Formation Wyoming, Part III. Eutheria and summary. *Univ. Calif. Publ. Geol. Sci.* **94**, 1–102 (1973).
21. Horovitz, I. Postcranial skeleton of *Ukhaatherium nessovi* (Eutheria, Mammalia) from the Late Cretaceous of Mongolia. *J. Vert. Paleontol.* **23**, 857–868 (2003).
22. Woodburne, M. O., Rich, T. A. & Springer, M. S. The evolution of tribospheny and the antiquity of mammalian clades. *Mol. Phylogen. Evol.* **28**, 360–385 (2003).
23. Rougier, G. W., Forasiepi, A. M., Martinelli, A. G. & Novacek, M. J. New Jurassic mammals from Patagonia, Argentina: a reappraisal of australosphenidan morphology and interrelationships. *Am. Mus. Novitates* **3566**, 1–54 (2007).
24. Asher, R. J., Novacek, M. J. & Geisler, J. H. Relationships of endemic African mammals and their fossil relatives based on morphological and molecular evidence. *J. Mammal. Evol.* **19**, 131–194 (2003).
25. Zack, S. P., Penkrot, T. A., Bloch, J. I. & Rose, K. D. Affinities of 'hyopsodontids' to elephant shrews and a Holarctic origin of Afrotheria. *Nature* **434**, 497–501 (2005).
26. Asher, R. J. et al. Stem Lagomorpha and the antiquity of Glires. *Science* **307**, 1091–1094 (2005).
27. Archibald, J. D. & Deutschman, D. H. Quantitative analysis of the timing of the origin and diversification of extant placentals. *J. Mammal. Evol.* **8**, 107–124 (2001).
28. Arnason, U. et al. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc. Natl. Acad. Sci. USA* **99**, 8151–8156 (2002).
29. Bininda-Emonds, O. R. P. et al. The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
30. Prasad, G. V. R., Jaeger, J. J., Sahni, A., Gheerbrant, E. & Khajuria, C. K. Eutherian mammals from the Upper Cretaceous (Maastrichtian) Intertrappean beds of Naskal, Andhra Pradesh, India. *J. Vert. Paleontol.* **14**, 260–277 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Davidson for specimen preparation; P. Bowden for illustration; N. B. Simmons, J. J. Flynn, R. L. Cifelli, J. G. Mead and H. L. Kafka for specimens; and J. D. Archibald, K. C. Beard, J. I. Bloch, R. L. Cifelli, M. R. Dawson, T. J. Gaudin, J. A. Hopson, I. Horovitz, Z. Kielan-Jaworowska, Z.-X. Luo, G. Metais, M. A. O'Leary, K. D. Rose and S. P. Zack for discussions. This work was supported by the NSF (J.R.W., G.W.R, M.J.N. and R.J.A.), the Carnegie Museum of Natural History, and the American Museum of Natural History.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R.W. (wiblej@carnegiemnh.org).

LETTERS

Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis

Trista E. North^{1,2}, Wolfram Goessling^{1,2}, Carl R. Walkley^{1,3}, Claudia Lengerke¹, Kamden R. Kopani^{1,2}, Allegra M. Lord^{1,2}, Gerhard J. Weber^{1,2}, Teresa V. Bowman^{1,2}, Il-Ho Jang¹, Tilo Grosser⁴, Garret A. FitzGerald⁴, George Q. Daley¹, Stuart H. Orkin^{1,2,3} & Leonard I. Zon^{1,2}

Haematopoietic stem cell (HSC) homeostasis is tightly controlled by growth factors, signalling molecules and transcription factors. Definitive HSCs derived during embryogenesis in the aorta–gonad–mesonephros region subsequently colonize fetal and adult haematopoietic organs^{1,2}. To identify new modulators of HSC formation and homeostasis, a panel of biologically active compounds was screened for effects on stem cell induction in the zebrafish aorta–gonad–mesonephros region. Here, we show that chemicals that enhance prostaglandin (PG) E2 synthesis increased HSC numbers, and those that block prostaglandin synthesis decreased stem cell numbers. The cyclooxygenases responsible for PGE2 synthesis were required for HSC formation. A stable derivative of PGE2 improved kidney marrow recovery following irradiation injury in the adult zebrafish. In murine embryonic stem cell differentiation assays, PGE2 caused amplification of multipotent progenitors. Furthermore, *ex vivo* exposure to stabilized PGE2 enhanced spleen colony forming units at day 12 post transplant and increased the frequency of long-term repopulating HSCs present in murine bone marrow after limiting dilution competitive transplantation. The conserved role for PGE2 in the regulation of vertebrate HSC homeostasis indicates that modulation of the prostaglandin pathway may facilitate expansion of HSC number for therapeutic purposes.

A chemical genetic screen was conducted to identify new pathways modulating definitive HSC formation during zebrafish embryogenesis. *runx1* and *cmyb*, required for mammalian HSC development, are expressed in the ventral wall of the dorsal aorta in a region analogous to the mammalian aorta–gonad–mesonephros (AGM) at 36 h post fertilization (h.p.f.)^{3–5}. Wild-type embryos, incubated with individual chemicals, were examined for alterations in *runx1*^{+/+}/*cmyb*^{+/+} HSCs by *in situ* hybridization expression at 36 h.p.f. A high percentage of compounds (91.7%, 2,275 of 2,357) failed to alter HSC expression, whereas 35 (1.4%) and 47 (1.9%) led to increased or decreased numbers of HSCs, respectively. Among these substances, 10 affected the prostaglandin pathway (Supplementary Table 1). *runx1*^{+/+}/*cmyb*^{+/+} HSCs comprise a line of flattened endothelial cells (arrow) and haematopoietic clusters (arrowhead) in the aorta (Fig. 1a–c); linoleic acid increased HSC numbers (22 altered out of 30 scored), whereas celecoxib, a cyclooxygenase (Cox)2 inhibitor, decreased HSCs (26/31). PGE2 is the main effector prostanoid produced in the zebrafish⁶ and is regulated by both Cox1 (also known as Ptgs1) and Cox2 (also known as Ptgs2a). Treatment of zebrafish embryos with PGE2 increased expression of *runx1*/*cmyb* (25/49), whereas Cox inhibition with SC560 (Cox1) and NS398 (Cox2) (Supplementary Fig. 1a–e) decreased HSC numbers in 30/36 and 35/44 cases, respectively.

These findings argue persuasively for a specific role of PGE2 in the formation of AGM HSCs.

Cox1 is required for the development of the aorta–vein endothelial boundary during zebrafish development⁷; thus, alteration in Cox1 activity could have an impact on endothelial-derived HSCs. By *in situ* hybridization, *cox2* was diffusely expressed in the tail region at 36 h.p.f. (Supplementary Fig. 1f, g). In FACS-isolated blood and endothelial cell populations, both *cox1* and *cox2* were found to be highly expressed during the onset of definitive haematopoiesis. *cox1* was detected in both *Lmo2*⁺ endothelial cells and in *Cd41*⁺ HSCs, whereas *cox2* was only found in HSCs (Supplementary Fig. 1h). This suggests that Cox1 and Cox2 participate in HSC induction through regulation of the stem cell niche and the HSC itself.

A long-acting derivative of PGE2, 16,16-dimethyl-PGE2 (dmPGE2) caused an increase in *runx1*^{+/+}/*cmyb*^{+/+} AGM HSCs in 78% of embryos (97/124) (Fig. 1e, h), whereas HSCs were inhibited by indomethacin (10 µM) treatment in 90% of embryos (92/102) (Fig. 1k and Supplementary Fig. 1j–r). Mass spectrometry of 36 h.p.f. embryos demonstrated that indomethacin treatment depressed PGE2 formation below detectable levels (from 18±6 pg per 50 embryos to <2 pg per 50 embryos; *n*=3)⁶. dmPGE2 had minimal effects on the vasculature, as shown by *flk1* staining (Fig. 1f, i) whereas indomethacin slightly altered the intersomitic vessels in 30% (15/49) of embryos (Fig. 1l). At 36 h.p.f., live bigenic *cmyb*–*gfp*; *lmo2*–*dsRed* (green-fluorescent-protein-labelled HSCs and progenitors; red-labelled HSCs and endothelium) embryos imaged by confocal microscopy exhibited significantly decreased numbers of HSCs (yellow) following indomethacin treatment, and significantly increased HSCs after dmPGE2 exposure (Fig. 1g, j, m and Supplementary Fig. 1i). Quantitative PCR confirmed an enhancement in *runx1* and *cmyb* expression by dmPGE2, whereas indomethacin significantly reduced the expression of each gene (Fig. 1d).

To confirm the requirement of PGE2 activity, we used morpholino oligonucleotides to knock down expression of Cox1 and Cox2; a low dose (40 µM) inhibition of Cox1 minimizes toxicity, while mimicking Cox-dependent developmental defects⁶. Morpholino oligonucleotide knockdown of Cox1/Cox2 decreased the levels of prostaglandins and inhibited AGM HSCs (Cox1, 54/74; Cox2, 60/71) (Supplementary Fig. 1s–u). The morpholino-mediated effects on HSCs were reversed by dmPGE2 (Cox1 + dmPGE2, 29/52 rescued; Cox2 + dmPGE2, 43/60) (Supplementary Fig. 1y, z, a'). dmPGE2 rescued (25/45) morpholino-mediated knockdown of PGE2 synthase (35/50) indicating that PGE2 signalling was sufficient to modulate HSC formation (Supplementary Fig. 1u, b'). PGE2 signals through receptors Ptger1l–Ptger4l (ref. 8). Morpholino-mediated

¹Stem Cell Program and Division of Hematology/Oncology, Children's Hospital, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA. ³Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁴Institute for Translational Medicine and Therapeutics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

knockdown of *Ptger2l* and *Ptger4l* diminished *runx1/cmyb* expression (*Ptger2l*, 39/63; *Ptger4l*, 44/67) and was not reversed by dmPGE2 (Supplementary Fig. 1s, t, c', d'). Quantitative PCR analysis showed *ptger2l/ptger4l* are present in HSCs (Supplementary Fig. 1e'). These experiments confirm that PGE2-mediated signalling regulates the formation of HSCs in the AGM region.

To examine the role of PGE2 in HSC homeostasis in adult zebrafish, we performed a kidney marrow irradiation-recovery assay in sublethally irradiated wild-type fish⁹ (Fig. 2a). The rate of kidney marrow repopulation was significantly enhanced after exposure to 50 μ M dmPGE2 (Fig. 2a, b), with progenitor recovery preceding reconstitution of the myeloid and lymphoid populations.

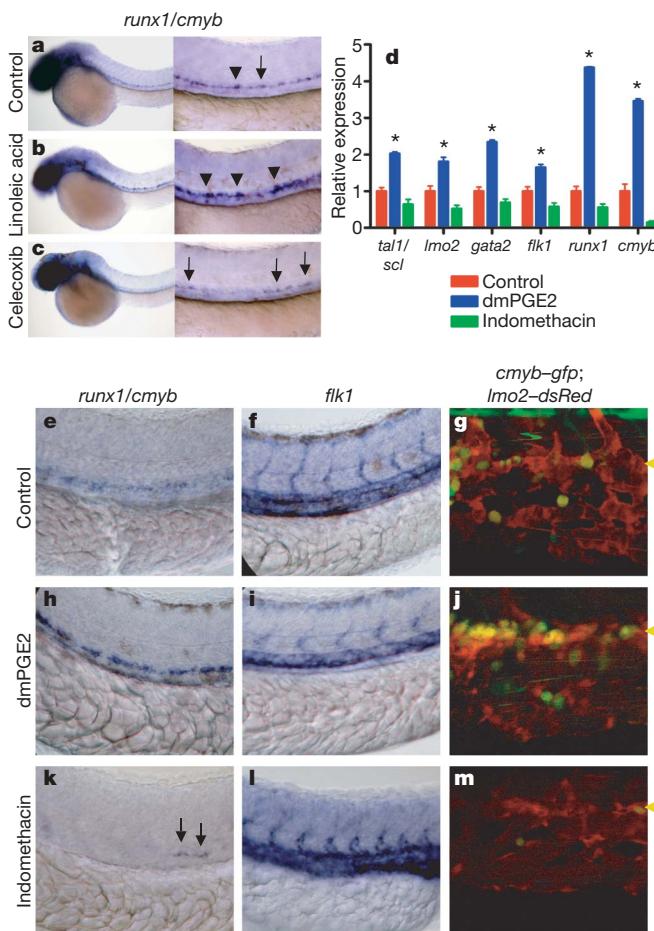


Figure 1 | Prostaglandin agonists and antagonists alter *runx1/cmyb* expression without affecting vascular development. *In situ* hybridization for *runx1/cmyb* or *flk1* at 36 h.p.f. Photomicrographs were taken with Nomarski optics at 10 \times (a–c, left panels) and 40 \times (a–c, right panels, and e–m) magnification. a–c, Representative examples of chemicals in the prostaglandin pathway discovered in the screen are shown; 10 μ M linoleic acid increases, and 20 μ M celecoxib reduces HSC numbers. *runx1*⁺/*myb*⁺ HSCs are indicated: endothelial cells (arrow); haematopoietic clusters (arrowhead). d, Quantitative PCR profile of endothelial and HSC-specific gene expression following exposure to long-acting dmPGE2 (10 μ M, blue) or the nonspecific Cox inhibitor indomethacin (10 μ M, green) versus control (red). Both treatments resulted in statistically significant differences compared with controls for each gene examined (ANOVA, * $P < 0.05$; mean, s.d. and n are listed in Supplementary Table 2). e–m, dmPGE2 and indomethacin exert opposing effects on *runx1/cmyb* expression by *in situ* hybridization (e, h, k); *flk1* is used to assess the effects on vascular development (f, i, l). Confocal microscopy images of *cmyb*-gfp; *lmo2*-dsRed bigenic fish exposed to dmPGE2 and indomethacin showing an increase and decrease in HSC (yellow) number along the ventral wall (yellow arrowhead) of the aorta, respectively (g, j, m). Quantitative analysis of 10 embryos in each treatment group revealed significant differences in HSC numbers (Supplementary Fig. 1i).

Significant upregulation of stem, progenitor and endothelial cell markers was found after dmPGE2 treatment (Supplementary Fig. 2a). Inhibition of Cox activity significantly decreased kidney marrow recovery and affected overall survival (Supplementary Fig. 2b). Our results indicate that PGE2 has an important role in kidney marrow homeostasis.

We then evaluated the effects of PGE2 on murine HSC and progenitor populations. Addition of dmPGE2 to embryonic stem cells during embryoid body expansion increased haematopoietic colonies on an OP9 stromal cell layer and in methylcellulose assays¹⁰ (Fig. 3a, b). OP9, definitive erythroid and granulocyte/monocyte colonies increased in a dose-dependent manner after exposure to 10 μ M (granulocyte/monocyte, $P = 0.005$) and 20 μ M (OP9, $P = 0.047$; definitive erythroid, $P = 0.04$; granulocyte/monocyte, $P = 0.007$) dmPGE2. The number of multipotent granulocyte/erythrocyte/monocyte/macrophage colonies was enhanced 2.9-fold following dmPGE2 treatment (10 μ M, $P = 0.017$; 20 μ M, $P = 0.016$). *Cox1* (also known as *Ptgs1*), *Cox2* (also known as *Ptgs2*), PGE2 synthase (*Ptges*) and *Ptger1*–*Ptger4* were present in embryonic stem cells at all stages examined (Supplementary Fig. 3a). Indomethacin inhibited colony growth at 20 μ M (OP9, $P = 0.069$) and 100 μ M (granulocyte/monocyte, $P = 0.024$) (Fig. 3a, b) and could be rescued by dmPGE2 (Supplementary Fig. 3b, c). These data suggest the role of PGE2 in regulating haematopoiesis is conserved between zebrafish and mammals.

To explore effects in an intact mammalian model, murine whole bone marrow (WBM) was exposed *ex vivo* to dmPGE2 (1 μ M per 10 6 cells) and irradiated recipients were transplanted with 6 \times 10 4 treated WBM cells. The number of spleen colony-forming units at day 12 post transplant (CFU-S12) was increased threefold ($P < 0.0001$) in recipients of dmPGE2-treated WBM (Fig. 4b, Supplementary Fig. 4b, Supplementary Table 6); numbers of more mature CFU-S8 colonies were also enhanced (Fig. 4a, Supplementary Fig. 3a, Supplementary Table 5). To assess the endogenous PGE2 requirement, WBM cells were incubated with indomethacin (1 μ M 10 $^{-6}$ cells) or specific COX1 and COX2 inhibitors. After transplantation of 1 \times 10 5 cells, a significant decrease ($P = 0.0001$) in the number of CFU-S12 was observed (Fig. 4c; Supplementary Fig. 4c, k, l; Supplementary Table 6). These results suggest that PGE2 enhances haematopoietic progenitor formation, and is required for CFU-S activity.

The prostaglandin pathway components are present in both stromal cell and HSC populations in mice and humans^{11,12}. *Cox1*, *Cox2*, PGE2-synthase, *Ptger2* and *Ptger4* are present in fetal liver HSCs and in bone marrow HSCs after 5-fluorouracil (5-FU) injury, suggesting PGE2 signalling functions in murine HSCs¹³. To determine if the increase in CFU-S number is due to a direct effect of PGE2 on the stem/progenitor cell population, FACS-isolated *cKit*⁺*Scal1*⁺ Lineage[–] (KSL) bone marrow cells were exposed to dmPGE2 and transplanted into irradiated recipients. Both splenic weight (Supplementary Fig. 4d) and CFU-S12 were significantly increased, indicating that dmPGE2 can lead to cell-autonomous activation of HSCs and immature progenitors (Fig. 4d, Supplementary Table 6).

A limiting dilution competitive repopulation analysis was conducted to determine the effects of dmPGE2 on HSC reconstitution¹⁴. WBM (CD45.1) exposed to dmPGE2 *ex vivo* was mixed independently at varying doses with a fixed number of untreated competitor cells (CD45.1/CD45.2) and injected into congenic recipient mice (CD45.2). Peripheral blood obtained at 6, 12 and 24 weeks post transplantation was examined by FACS to determine percentage test-cell contribution to haematopoietic repopulation (Supplementary Fig. 4e–j); positive reconstitution was defined as test-cell multilineage chimaerism $>5\%$ (Supplementary Fig. 4f, h, i). A significant increase in the number of repopulating cells as determined by Poisson statistics was seen in dmPGE2-treated bone marrow (Fig. 4e, Supplementary Fig. 4g, j). At 6 weeks, the calculated frequency of engrafting cells per 10 6 WBM cells was enhanced 3.3-fold ($P = 0.005$) in dmPGE2-treated WBM recipients, and the frequency of

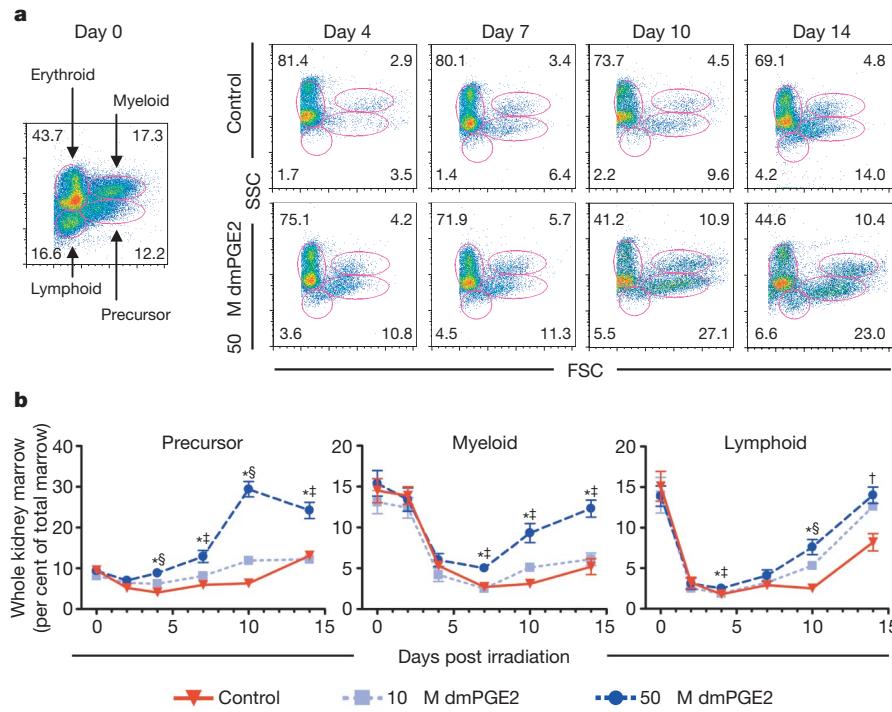


Figure 2 | Treatment with dmPGE2 enhances haematopoietic recovery in sublethally irradiated adult zebrafish. Zebrafish whole kidney marrow irradiation recovery experiments were performed. **a**, Representative FSC/SSC FACS profiles of haematopoietic cell lineages in the kidney marrow on days 0, 4, 7, 10 and 14 of irradiation recovery in DMSO and dmPGE2-treated

(50 μ M) zebrafish. **b**, Kinetics of kidney marrow reconstitution of precursor, lymphoid and myeloid cells in control and dmPGE2-treated fish. Statistically significant differences: \dagger , 50 μ M versus control; \ddagger , 50 μ M versus 10 μ M, and 50 μ M versus control; and \S , all variables significant (ANOVA, $*P < 0.05$; mean, s.d. and n listed in Supplementary Table 3).

short-term repopulating HSCs was 4.0-fold ($P = 0.002$) higher at 12-weeks post-transplantation (Fig. 4e, f, Supplementary Fig. 4g). At 24 weeks, the frequency of long-term repopulating HSCs was 2.3-fold enhanced ($P = 0.05$) in recipients of dmPGE2-treated cells (Fig. 4f; Supplementary Fig. 4j). dmPGE2 treatment increased the frequency of repopulating HSCs in the mouse, but did not impair the differentiative capacity as seen by multilineage analysis. To determine whether dmPGE2 treatment enhanced homing to the bone marrow niche, WBM was labelled with a vital dye, CFDA, before transplantation; no significant difference in homing could be detected ($P = 0.83$) (Supplementary Fig. 5).

Here we have demonstrated that PGE2 enhances the number of HSCs and multipotent progenitors in two vertebrate species, zebrafish and mice. Prior studies have documented that unmodified PGE2 impairs blood-cell maturation in the mouse^{15,16} and cell cycle stimulation in CFU-S8 progenitors¹⁷; however, the effects of prostaglandin-mediated cell signalling on HSCs have not been examined previously. Cox1 and Cox2 seem to have distinct functions in AGM HSC formation: Cox1 is important in the formation of the haematopoietic niche, whereas Cox2 is probably involved in self-renewal and proliferation of HSCs themselves. Conversely, homozygous *Cox1* or *Cox2* knockout mice are viable, without apparent defects in HSC formation¹⁸, due to

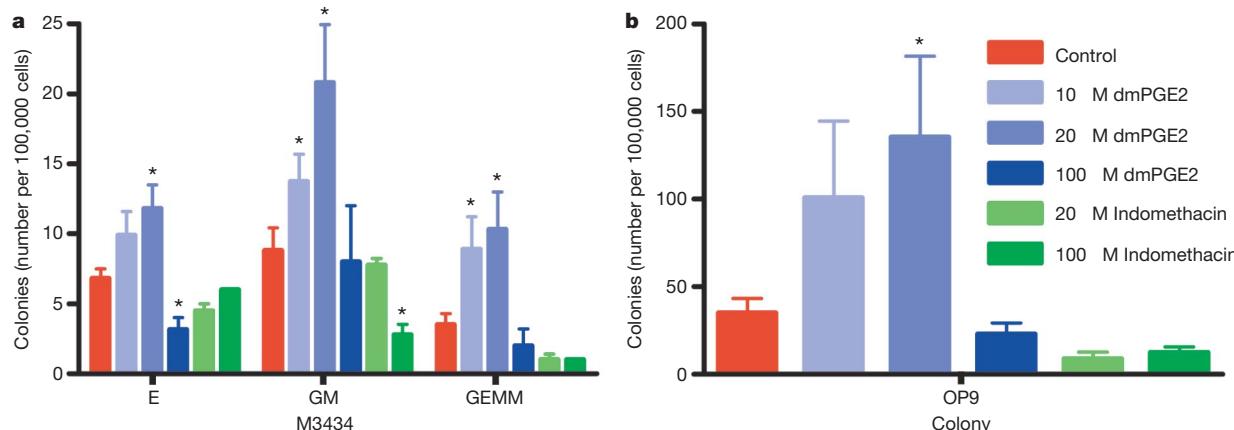


Figure 3 | dmPGE2 modulates colony number and haematopoietic differentiation in mouse embryonic stem cells. M3434 and OP9 embryonic stem cell colony-forming assays were performed; counts are per 100,000 cells plated. An asterisk (*) indicates a statistically significant difference (two-tailed *t*-test; mean, s.d. and n listed in Supplementary Table 4). **a**, Effect of increasing doses of dmPGE2 and inhibition of cyclooxygenase activity by indomethacin on haematopoietic differentiation in methylcellulose;

numbers of definitive erythroid (E), mixed granulocyte/monocyte (GM), and multi-potent (GEMM) progenitor colonies are shown (10 μ M dmPGE2: GM, $P = 0.005$; GEMM, $P = 0.017$; 20 μ M dmPGE2: E, $P = 0.04$; GM, $P = 0.007$; GEMM, $P = 0.016$; 100 μ M indomethacin: GM, $P = 0.024$). **b**, Effect of dmPGE2 and indomethacin on OP9 haematopoietic colony number (20 μ M dmPGE2, $P = 0.047$).

maternal and sibling PGE2 contribution^{7,19}. Analyses of *Cox2*^{-/-} mice demonstrated alterations in haematocrit levels and an impaired recovery from 5-FU-induced bone marrow injury²⁰; these findings imply HSC defects in adult *Cox2*^{-/-} mice that are compatible with our proposed role for prostaglandin in HSC homeostasis. To clarify the roles of COX1 and COX2 in regulating HSC homeostasis, we performed CFU-S12 (Fig. 4k, l) and 5-FU bone marrow recovery assays using selective inhibitors of COX1 (SC560) or COX2 (NS398). Inhibition of either enzyme significantly diminished CFU-S activity, as well as the recovery of peripheral blood and bone marrow WBC numbers (Supplementary Fig. 4m, n). Administration of dmPGE2 following 5-FU treatment significantly enhanced bone marrow recovery. These data suggest that both COX1 and COX2 have a role in regulating HSC homeostasis in adult mice, as in the zebrafish, and that PGE2 is the mediator of this HSC regulation. The precise mechanism of PGE2 modulation of vertebrate HSC homeostasis remains to be elucidated.

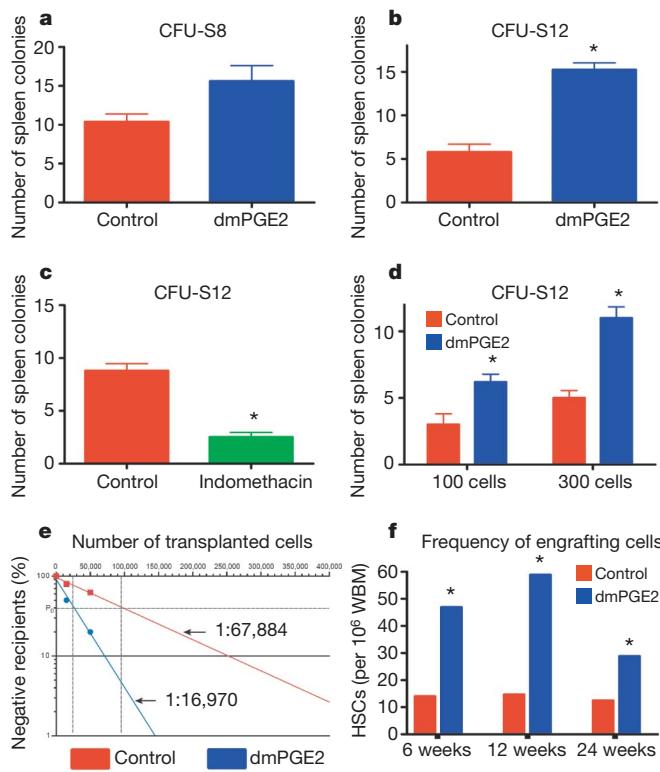


Figure 4 | Exposure of murine bone marrow to dmPGE2 increases the number of CFU-S and repopulating HSCs. An asterisk (*) indicates a statistically significant difference. **a, b,** Effect of *ex vivo* treatment of WBM (2 h on ice) with ethanol control (red) or dmPGE2 (1 μ M per 10^6 cells) on CFU-S8 and CFU-S12 (60,000 cells per recipient; CFU-S12: two-tailed *t*-test, control (mean/s.d./n) = 5.78/2.73/9, dmPGE2 = 15.22/2.39/9, $P < 0.0001$). **c,** Effect on CFU-S12 following *ex vivo* treatment with indomethacin (1 μ M per 10^6 cells) (100,000 cells/recipient; two-tailed *t*-test, control (mean/s.d./n) = 8.8/2.10/10, indomethacin = 2.5/1.43/10, $P = 0.0001$). **d,** CFU-S12 evaluation after treatment of cKit⁺Sca1⁺Lineage⁻ stem cells with dmPGE2 or ethanol control (two-tailed *t*-test, 100 cells: control (mean/s.d./n) = 3/1.63/4, dmPGE2 = 6.2/1.3/5, $P = 0.013$; 300 cells: control (mean/s.d./n) = 5/1.22/5, dmPGE2 = 11/1.87/5, $P = 0.0003$). **e, f,** Limiting dilution competitive repopulation assay. The number of negative recipients as determined by FACS analysis (e) in relation to the total number of cells transplanted for control or dmPGE2-treated cell samples is shown at 12 weeks. $P_0 = 67,884$ (control) and 16,970 (dmPGE2 treated). The frequency of engraftment (f) at 6, 12, and 24 weeks post transplantation in recipients of ethanol- versus dmPGE2-treated WBM calculated by Poisson statistics (ANOVA, $n = 10$ per variable; 6 wks, $P = 0.005$; 12 wks, $P = 0.002$; 24 wks, $P = 0.05$); the number of recipients surviving to analysis at each time point is shown in Supplementary Tables 7–9.

Patients undergoing bone marrow transplantation show increased endogenous PGE2 levels²¹. Our studies raise the possibility that administration of COX inhibitors following human bone marrow transplantation might impair HSC engraftment and result in delayed recovery of the WBC counts, in addition to adversely affecting platelet function. PGE2 and its analogues are safely administered to patients^{22,23}. The concentration of dmPGE2 used to expand the number of murine HSCs falls within the physiological range of PGE2 in human serum²⁴, thus dmPGE2 or its derivatives may be useful for *ex vivo* or *in vivo* expansion of HSCs. Our studies illustrate that PGE2 functions as a potent regulator of HSCs in vertebrates, and may prove useful in treating patients with bone marrow failure or following transplantation.

METHODS SUMMARY

Wild-type age-matched embryos were exposed to individual test compounds from 3-somites until 36 h.p.f. and effects on HSCs were evaluated by *in situ* hybridization for *runx1* and *cmyb*. Treatment with PGE2, dmPGE2 and Cox inhibitors (indomethacin, SC560, NS398) at 10 μ M was used to confirm and quantify the effects of prostaglandin signalling on HSCs by *in situ* hybridization, quantitative PCR⁹ and confocal microscopy²⁵. Expression of prostaglandin pathway components in HSCs was characterized by microarray analysis²⁶, quantitative PCR and *in situ* hybridization. Morpholino knockdown of prostaglandin pathway components^{6,8,27}, and subsequent rescue by dmPGE2, was used to confirm the specificity of the results of the chemical treatments. Functional inhibition of prostaglandin synthesis was measured by mass spectroscopy for chemical and morpholino experiments. A flow-cytometry-based irradiation recovery assay was used to assess the impact of PGE2-mediated signalling on adult kidney marrow²⁸. The effect of dmPGE2 and indomethacin on haematopoietic colony forming potential of embryonic stem cells was analysed by standard OP9 and methylcellulose colony forming assays^{29,30}. CFU-S assays and limiting dilution competitive transplantation assays were used to test the effects of *ex vivo* dmPGE2 treatment or COX inhibition on haematopoietic stem and progenitor populations. Bone marrow ablation¹³ by 5-FU was used to test the *in vivo* effect of dmPGE2 or COX inhibitor treatment on haematopoietic recovery in mammals.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 December 2006; accepted 30 April 2007.

- Dzierzak, E. The emergence of definitive hematopoietic stem cells in the mammal. *Curr. Opin. Hematol.* 12, 197–202 (2005).
- Galloway, J. L. & Zon, L. I. Ontogeny of hematopoiesis: examining the emergence of hematopoietic cells in the vertebrate embryo. *Curr. Top. Dev. Biol.* 53, 139–158 (2003).
- North, T. E. et al. Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* 16, 661–672 (2002).
- Mukouyama, Y. et al. Hematopoietic cells in cultures of the murine embryonic aorta-gonad-mesonephros region are induced by c-Myb. *Curr. Biol.* 9, 833–836 (1999).
- Kalev-Zylinska, M. L. et al. Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. *Development* 129, 2015–2030 (2002).
- Grosser, T., Yusuff, S., Cheskis, E., Pack, M. A. & FitzGerald, G. A. Developmental expression of functional cyclooxygenases in zebrafish. *Proc. Natl. Acad. Sci. USA* 99, 8418–8423 (2002).
- Cha, Y. I., Kim, S. H., Solnica-Krezel, L. & Dubois, R. N. Cyclooxygenase-1 signaling is required for vascular tube formation during development. *Dev. Biol.* 282, 274–283 (2005).
- Cha, Y. I. et al. Cyclooxygenase-1-derived PGE2 promotes cell motility via the G-protein-coupled EP4 receptor during vertebrate gastrulation. *Genes Dev.* 20, 77–86 (2006).
- Burns, C. E., Traver, D., Mayhall, E., Shepard, J. L. & Zon, L. I. Hematopoietic stem cell fate is established by the Notch-Runx pathway. *Genes Dev.* 19, 2331–2342 (2005).
- Nakano, T., Kodama, H. & Honjo, T. *In vitro* development of primitive and definitive erythrocytes from different precursors. *Science* 272, 722–724 (1996).

11. Ivanova, N. B. et al. A stem cell molecular signature. *Science* **298**, 601–604 (2002).
12. Akashi, K. et al. Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101**, 383–389 (2003).
13. Venezia, T. A. et al. Molecular signatures of proliferation and quiescence in hematopoietic stem cells. *PLoS Biol.* **2**, e301 (2004).
14. Zhang, C. C. & Lodish, H. F. Insulin-like growth factor 2 expressed in a novel fetal liver cell population is a growth factor for hematopoietic stem cells. *Blood* **103**, 2513–2521 (2004).
15. Boer, A. K., Drayer, A. L., Rui, H. & Vellenga, E. Prostaglandin-E2 enhances EPO-mediated STAT5 transcriptional activity by serine phosphorylation of CREB. *Blood* **100**, 467–473 (2002).
16. Rocca, B. et al. Cyclooxygenase-2 expression is induced during human megakaryopoiesis and characterizes newly formed platelets. *Proc. Natl Acad. Sci. USA* **99**, 7634–7639 (2002).
17. Feher, I. & Gidali, J. Prostaglandin E2 as stimulator of haemopoietic stem cell proliferation. *Nature* **247**, 550–551 (1974).
18. Langenbach, R., Loftin, C., Lee, C. & Tiano, H. Cyclooxygenase knockout mice: models for elucidating isoform-specific functions. *Biochem. Pharmacol.* **58**, 1237–1246 (1999).
19. Langenbach, R. et al. Prostaglandin synthase 1 gene disruption in mice reduces arachidonic acid-induced inflammation and indomethacin-induced gastric ulceration. *Cell* **83**, 483–492 (1995).
20. Lorenz, M. et al. Cyclooxygenase-2 is essential for normal recovery from 5-fluorouracil-induced myelotoxicity in mice. *Exp. Hematol.* **27**, 1494–1502 (1999).
21. Cayeux, S. J., Beverley, P. C., Schulz, R. & Dorken, B. Elevated plasma prostaglandin E2 levels found in 14 patients undergoing autologous bone marrow or stem cell transplantation. *Bone Marrow Transplant.* **12**, 603–608 (1993).
22. Talosi, G. et al. Prostaglandin E1 treatment in patent ductus arteriosus dependent congenital heart defects. *J. Perinat. Med.* **32**, 368–374 (2004).
23. Thanopoulos, B. D., Andreou, A. & Frimas, C. Prostaglandin E2 administration in infants with ductus-dependent cyanotic congenital heart disease. *Eur. J. Pediatr.* **146**, 279–282 (1987).
24. Hertelendy, F., Woods, R. & Jaffe, B. M. Prostaglandin E levels in peripheral blood during labor. *Prostaglandins* **3**, 223–227 (1973).
25. Zhu, H. et al. Regulation of the *lmo2* promoter during hematopoietic and vascular development in zebrafish. *Dev. Biol.* **281**, 256–269 (2005).
26. Weber, G. J. et al. Mutant-specific gene programs in the zebrafish. *Blood* **106**, 521–530 (2005).
27. Pini, B. et al. Prostaglandin E synthases in zebrafish. *Arterioscler. Thromb. Vasc. Biol.* **25**, 315–320 (2005).
28. Traver, D. et al. Effects of lethal irradiation in zebrafish and rescue by hematopoietic cell transplantation. *Blood* **104**, 1298–1305 (2004).
29. Kyba, M. et al. Enhanced hematopoietic differentiation of embryonic stem cells conditionally expressing Stat5. *Proc. Natl Acad. Sci. USA* **100** (Suppl 1), 11904–11910 (2003).
30. Wang, Y., Yates, F., Naveiras, O., Ernst, P. & Daley, G. Q. Embryonic stem cell-derived hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **102**, 19081–19086 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Institute of Chemical and Cellular Biology at Harvard Medical School for access to the chemical libraries used in the screen. We thank: A. Flint, E. Mayhall and C.E. Burns for technical assistance and advice on the kidney marrow analysis; C. Thisse and B. Thisse for information and plasmids for PTGER2; and A. Meyers and J. Ojeda for technical help with the zebrafish chemical screen. This work was supported by grants from the National Institutes of Health (T.E.N., W.G., G.Q.D., S.H.O., G.A.F. and L.I.Z.), the American Cancer Society (T.E.N.), the American Gastroenterological Association (W.G.), the Leukemia and Lymphoma Society (C.R.W.), the American Heart Association (T.G.) and the Dr. Mildred Scheel Foundation for Cancer Research (C.L.). S.H.O. and L.I.Z. are Howard Hughes Medical Institute investigators.

Author Contributions T.E.N. and K.R.K. conducted the chemical screen. T.E.N., W.G. and A.M.L. performed the zebrafish prostaglandin studies. T.E.N. and C.R.W. conducted the murine experiments. G.J.W. completed the microarray analysis. T.E.N., W.G. and T.V.B. performed 5-FU treatment. T.G. provided cox1 and cox2 probes and completed the mass spectroscopy analysis. C.L. and I.H.J. performed the embryonic stem cell assays. T.E.N., W.G. and L.I.Z. wrote the manuscript. All authors discussed results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.I.Z. (zon@enders.tch.harvard.edu).

METHODS

Chemical screen design and confirmatory testing. Wild-type age-matched embryos were arrayed into 48-well plates (~5 embryos per well) of individual test compounds and exposed from 3-somites until 36 h.p.f. Three compound libraries were used: NINDS Custom Collection (1,040 compounds), SpecPlus Collection (960) and BIOMOL ICCB Known Bioactives (480). Five per cent (123/2480) of the compounds were toxic, resulting in death or severe morphological abnormalities. *In situ* hybridization for *runx1* and *cmyb* was performed to assess HSC numbers. Compounds were retested at 10, 20 and 50 μM . Stem cell specificity was assessed using *flk1* at 36 h.p.f. PGE2, PGI2, dmPGE2 and all Cox inhibitors (Sigma) were used at 10 to 20 μM .

Qualitative scoring (number of embryos with altered HSCs per number scored) of *runx1/cmyb* was conducted using the following criteria: Normal/unchanged, continuous line of *runx1⁺/cmyb⁺* endothelial cells and occasional haematopoietic clusters; decreased/absent, reduction in *runx1⁺/cmyb⁺* cells, including the presence of large gaps in the line of HSCs, isolated positive cells, or absence of expression; increased/excess, enhancement in *runx1⁺/cmyb⁺* cells, including many HSC clusters, a thickened line of HSCs, or ectopic expression. **Confocal imaging.** Live 36 h.p.f. treated bigenic zebrafish embryos were embedded in 1% low-melting point agarose containing 0.04 mg ml⁻¹ Tricaine-S for confocal imaging. *lmo2-dsRed* fish were created as described²⁵. *cmyb-gfp* transgenic reporter lines were created by homologous recombination of a 3.7 kb EGFP construct downstream of the 5' untranslated region and precisely before the start site of a PAC clone containing *cmyb* (J. Galloway, H. Zhu, S. Lin and L.I.Z., unpublished data). For HSC quantification, cMyb⁺/Lmo2⁺ cells were counted in projections of z-stack images ($n = 10$ per treatment).

Morpholino knockdown. Morpholino oligonucleotides (GeneTools) directed against zebrafish *cox1* and *cox2*, PGE2 synthase, and *ptger2l* and *ptger4l* (refs 6, 8, 27) were injected (40 μM) into zebrafish embryos at the one-cell stage. For rescue experiments, 3-somite-stage morpholino-injected embryos were exposed to 10 μM dmPGE2.

Microarray gene expression profiling. Gata1-Gfp⁺ (12 somites), Lmo2-Gfp⁺ (12 somites and 35 h.p.f.) and Cd41-Gfp⁺ (35 h.p.f.) cells were FACS-sorted; total RNA was purified and analysed using Affymetrix zebrafish gene chips as described previously²⁶.

Quantitative PCR. Quantitative PCR (qPCR) was performed using previously described primer sets⁹. Embryos ($n = 50$) were treated as described. qPCR (60 °C annealing) was performed using SYBR Green Supermix on the iQ5 Multicolour RTPCR Detection System (BioRad) ($n = 10$ replicates) and relative expression levels were determined. Primer pairs for *Ptger2* and *Ptger4* are shown (Supplementary Table 10). qPCR of whole kidney marrow RNA ($n = 15$ per variable) was performed on day 3 post irradiation as described. qPCR on S cell RNA (harvested in Stat-60, Tel-Test) was performed using the Stratagene Sybrgreen kit on the Stratagene qPCR machine. Prostaglandin pathway component primer sequences are shown (Supplementary Table 10).

Mass spectroscopy. PGE₂ and the stable PGI₂ metabolite, 6-keto-PGF_{1 α} , were measured using high-performance liquid chromatography–tandem mass spectrometry. Ethylacetate extracts from homogenized embryos were spiked with the corresponding stable-isotope-labelled internal standards (*d*₄-PGE₂ and *d*₄-6-keto PGF_{1 α}) and allowed to react with methoxylamine. The following mass transitions were monitored: *m/z* 384→272 (PGE), *m/z* 398→368 (6-keto PGF_{1 α} and TxB₂).

Irradiation recovery assay. Adult zebrafish were exposed to 23 Gy of γ -irradiation. On day 2 post irradiation, fish were exposed overnight to DMSO control, dmPGE2 (10 or 50 μM), indomethacin (10 μM), SC560 (10 μM) or NS398 (10 μM) in fish water. Whole kidney marrow isolated on days 0, 2, 4, 7, 10, 14 was subjected to forward scatter/side scatter (FSC/SSC) FACS analysis to identify haematopoietic lineages ($n = 5$ per treatment, 3 replicates)²⁸.

Embryonic stem cell differentiation assays. Embryonic stem cell haematopoietic differentiation assays were performed as previously described^{29,30}. dmPGE2 (10, 20 or 100 μM) or indomethacin (20, 100 μM) were added at day 4 and 5 during embryoid body expansion. M3434 methylcellulose colony forming and OP9 colony assays were conducted on day 6 and analysed at days 8 and 5, respectively. Colony type was identified by morphological analysis; duplicate chemical exposures were averaged to determine the reported colony number ($n = 3$ replicates minimum).

Murine colony-forming units-spleen (CFU-S). WBM cells from the femurs of 8-week-old C57Bl/6 mice were incubated *ex vivo* with (1 μM per 10⁶ cells) dmPGE2, indomethacin, SC560, NS398 or ethanol control on ice for 2 h. Two independent bone marrow samples were treated ($n = 5$ per treatment, 2 replicates) for each variable. Recipient mice were lethally irradiated with a split dose of 10 Gy. Sixty thousand unfractionated dmPGE2 or control-treated bone marrow cells were injected retro-orbitally into irradiated recipient mice. Spleens were dissected on day 8 or 12, weighed and fixed with Bouin's solution;

haematopoietic colonies per spleen were counted. Cells (1 × 10⁵ per recipient) were transplanted after treatment with the COX inhibitors. FACS-sorted cKit⁺Sca1⁺Lineage⁻ bone marrow cells were treated as above and transplanted at a dose of 100 or 300 cells per recipient.

5-fluorouracil bone marrow injury. Mice were treated with 5-FU (150 mg kg⁻¹) as described¹³. SC560, NS398, dmPGE2 (1 mg kg⁻¹) or ethanol control were administered by intraperitoneal injection on days 1, 5, 9, 13 and 17 post injection. Peripheral blood was obtained on day 7 and 14, quantified and subjected to multilineage FACS analysis using antibodies (eBioscience) to B220/IgM (B-lymphoid), CD4/8 (T-lymphoid), Mac1/Gr1 (myeloid), Ter119/CD71 (erythroid) and cKit/Sca1 (stem/progenitor). Mice were killed on day 16, and bone marrow was isolated, quantified and analysed by FACS.

Limiting dilution competitive transplantation. WBM from CD45.1 C57Bl/6 mice was incubated with dmPGE2 or ethanol control *ex vivo*, as described. Treated test cells were independently transplanted into irradiated CD45.2 recipients ($n = 5$ per variable, 2 replicates) with untreated CD45.1/CD45.2 competitor at the following ratios: 15,000:200,000 (0.075:1), 50,000:200,000 (0.25:1), 200,000:200,000 (1:1) or 2,000,000:200,000 (10:1). Peripheral blood was obtained at 6, 12 and 24 weeks post transplantation, and white blood cells were FACS-analysed to determine test reconstitution for each series of treatment populations. Frequency of peripheral blood chimaerism >5% was used to calculate the number of repopulating cells using the L-Calc program (Stem Cell Technologies). For 12- and 24-week peripheral blood samples, multilineage reconstitution was measured by FACS analysis as above.

LETTERS

Akt/PKB regulates hepatic metabolism by directly inhibiting PGC-1 α transcription coactivator

Xinghai Li¹, Bobby Monks¹, Qingyuan Ge² & Morris J. Birnbaum¹

Type 2 diabetes mellitus, a disease with significant effects on the health and economy of Western societies, involves disturbances in both lipid and carbohydrate metabolism^{1–3}. In the insulin-resistant or diabetic state, the liver is unresponsive to the actions of insulin with regard to the suppression of glucose output but continues to produce large amounts of lipid, the latter mimicking the fed, insulin-replete condition^{4,5}. The disordered distribution of lipids contributes to the cardiovascular disease that is the greatest cause of mortality of type 2 diabetes mellitus^{6,7}. Yet the precise signal transduction pathways by which insulin regulates hepatic lipid synthesis and degradation remain largely unknown. Here we describe a mechanism by which insulin, through the intermediary protein kinase Akt2/protein kinase B (PKB)- β , elicits the phosphorylation and inhibition of the transcriptional coactivator peroxisome proliferator-activated receptor-coactivator 1 α (PGC-1 α), a global regulator of hepatic metabolism during fasting. Phosphorylation prevents the recruitment of PGC-1 α to the cognate promoters, impairing its ability to promote gluconeogenesis and fatty acid oxidation. These results define a mechanism by which insulin controls lipid catabolism in the liver and suggest a novel site for therapy in type 2 diabetes mellitus.

Consistent with previously published data, insulin decreased the level of *G6Pase* and *PEPCK* messenger RNAs, transcripts of two genes (encoding glucose-6-phosphatase and phosphoenolpyruvate carboxykinase) that are important to gluconeogenesis and regulated cooperatively by PGC-1 α and FoxO1, a winged helix transcription factor; an inhibitor of Akt1 and Akt2 blocked the effects of insulin^{8–11} (Supplementary Fig. S1a). Akt/PKB is an obligate mediator of many, if not all of the metabolic actions of insulin^{12–15}. The ability of insulin to suppress the hepatic production of glucose depends on the phosphorylation of FoxO1 by Akt and its consequent inhibition (refs 9, 16, 17). Insulin also reduced, in an Akt-dependent manner, the level of mRNAs encoding the small heterodimer partner (SHP) and medium-chain acetyl-CoA dehydrogenase (MCAD), proteins involved in bile acid synthesis and fatty acid oxidation, respectively, but not regulated by FoxO1 (Supplementary Figs S1a and S2). The Akt inhibitor increased the mRNA levels in the absence of insulin, which is indicative of the high basal activity of this pathway in H4IIE hepatoma cells. Because PGC-1 α co-activates the transcription of *SHP* and *MCAD* as well as that of *PEPCK* and *G6Pase*, we sought a mechanism by which Akt could influence the actions of PGC-1 α independently of FoxO1 (refs 18–23).

The carboxy terminus of PGC-1 α contains multiple potential sites of Akt phosphorylation (RXRXXS/T), but only Ser 570 is followed by the preferred bulky hydrophobic residue phenylalanine (Supplementary Fig. S1b)²⁴. Moreover, the sequence surrounding Ser 570 is conserved across various species. Recombinant Akt phosphorylated a C-terminal fragment of glutathione S-transferase (GST)-conjugated

PGC-1 α , but not the amino-terminal fragment, or GST itself (Fig. 1a). To map the site(s) in more detail, a series of the GST–PGC-1 α C-terminal mutations of potential Akt sites was generated and tested in the *in vitro* kinase assay. As shown in Fig. 1a, mutation of Ser 570 to Ala markedly decreased the phosphorylation; mutation of Ser 570 in combination with the other putative Akt sites did not further decrease the extent of phosphorylation, suggesting that Ser 570 is the major phosphorylation site for Akt *in vitro*.

To explore whether PGC-1 α is phosphorylated in cells, a phospho-PGC-1 α antibody was generated and shown to be specific for the phosphorylated form of Ser 570 (Supplementary Fig. S3). H4IIE cells infected with adenoviral vectors expressing PGC-1 α or PGC-1 α with Ser 570 mutated to Ala (S570A) were treated with insulin or insulin plus LY294002, an inhibitor of phosphatidylinositol-3-OH kinase, and the cell lysates were probed with the phospho-specific antibody. As illustrated in Fig. 1b, insulin significantly increased the phosphorylation of PGC-1 α at Ser 570, and this was blocked by LY294002. To provide further support for the idea that the pPGC-1 α S570 antibody-recognized signal was caused by phosphorylation of PGC-1 α , H4IIE cells were infected with a recombinant adenovirus expressing Myc-tagged PGC-1 α and treated with insulin or insulin plus LY294002; immunoprecipitation was performed with an anti-Myc antibody. As shown in Fig. 1c, insulin significantly increased PGC-1 α phosphorylation and, as above, this phosphorylation was blocked by LY294002. These results indicate that treatment with insulin leads to the phosphorylation of PGC-1 α at Ser 570.

The western blot shown as Fig. 1b revealed that, even in the absence of adenoviral PGC-1 α , insulin stimulated the appearance of a faint band immunoreactive with the anti-pPGC-1 α S570, suggesting that insulin is capable of promoting the phosphorylation of endogenous PGC-1 α . To examine this more directly, cell lysates prepared from H4IIE cells treated with insulin were immunoprecipitated with the anti-pPGC-1 α S570 antibody, followed by western blotting with the anti-PGC-1 α antibody. As shown in Fig. 1d, about eightfold more phospho-PGC-1 α was precipitated after treatment with insulin, indicating enhanced phosphorylation of the endogenous protein in hepatoma cells. Infection of H4IIE cells with an adenovirus expressing a constitutively active mutant of Akt (MyrAkt) led to phosphorylation of PGC-1 α at Ser 570 to an extent comparable to that produced by insulin. Furthermore, insulin stimulated the phosphorylation of adenoviral expressed PGC-1 α in wild-type primary hepatocytes in a manner that was largely dependent on the presence of Akt2 (Fig. 1e). Last, insulin of injection into mice led to approximately fourfold more phospho-PGC-1 α being precipitated from liver nuclear extracts under conditions in which total PGC-1 α did not change (Fig. 1f). Insulin-stimulated PGC-1 α phosphorylation was blunted in the liver nuclear extracts from *Akt2*^{-/-} mice (Fig. 1f). These results suggest strongly that insulin promotes PGC-1 α phosphorylation at Ser 570 by an Akt-mediated pathway.

¹Institute for Diabetes, Obesity and Metabolism, Cox Institute, University of Pennsylvania School of Medicine and the Howard Hughes Medical Institute, Philadelphia, Pennsylvania 19104, USA. ²Cell Signaling Technology, Inc., 166B Cummings Center, Danvers, Massachusetts 01923, USA.

MyrAkt reduced the levels of the mRNAs encoding G6Pase and PEPCK, two key enzymes in gluconeogenesis (Fig. 2a). Overexpression of PGC-1 α increased *G6Pase* and *PEPCK* mRNA levels 20-fold and 6-fold, respectively, and coexpression of MyrAkt significantly abrogated this induction. If it is correct that insulin antagonizes the actions of PGC-1 α by Akt-dependent phosphorylation, Akt should be impaired in its ability to counter the effects of S570A. As shown in Fig. 2a, both PGC-1 α and S570A elevated the *G6Pase* and *PEPCK* mRNA levels to equivalent extents, but coexpression of MyrAkt was unable to suppress S570A-activated transcription of these genes. To address the importance of this mechanism *in vivo* with a physiologically

meaningful endpoint, we examined the ability of Akt to antagonize the induction of glucose production by PGC-1 α or S570A in primary hepatocytes⁹. As shown in Fig. 2b, both PGC-1 α and S570A increased glucose production about sixfold; coexpression of MyrAkt significantly antagonized PGC-1 α -activated glucose production, but suppressed S570A-mediated glucose production to a smaller extent. The incomplete reduction in the latter condition is probably due to the contribution of other Akt targets such as FoxO1 phosphorylation in the inhibition of gluconeogenesis⁹. Nonetheless, these data show that complete suppression of glucose output from liver cells depends on the ability of Akt to phosphorylate PGC-1 α at Ser 570.

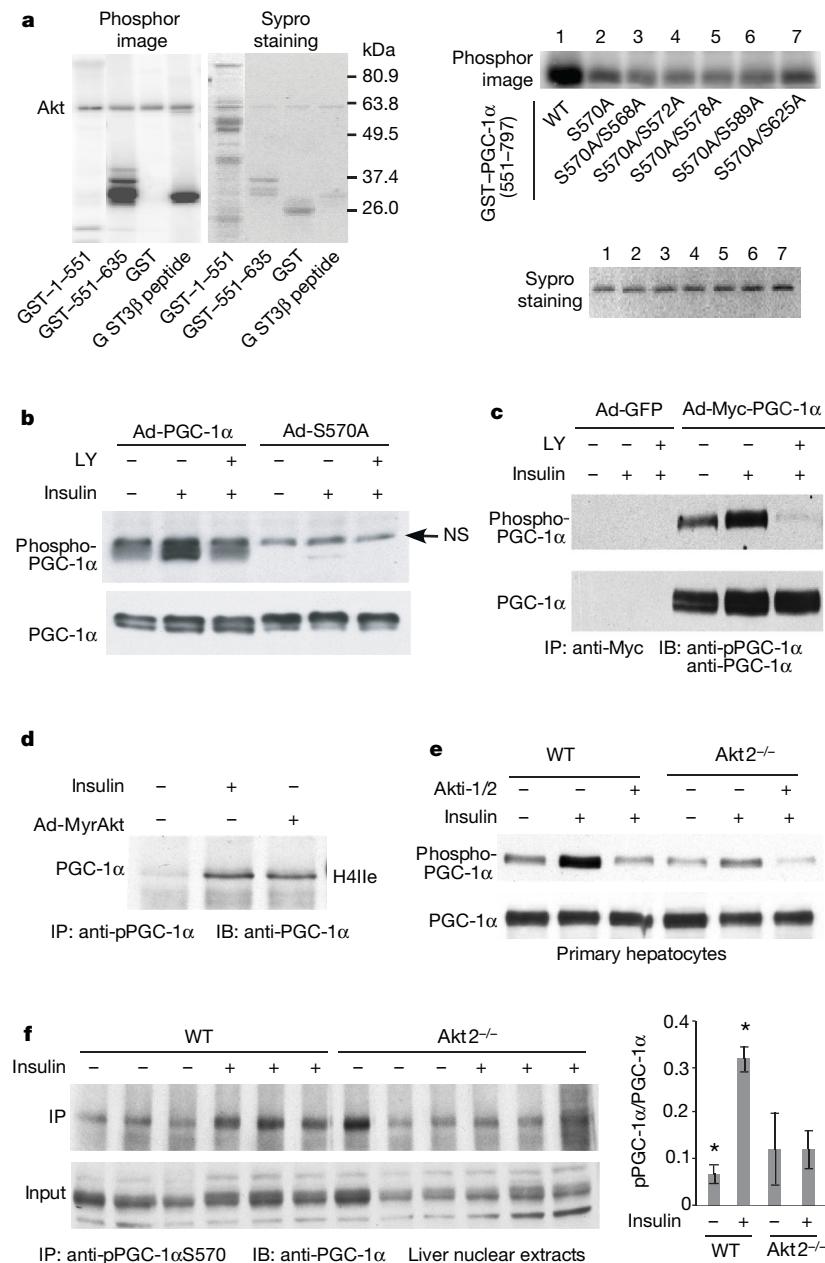


Figure 1 | Akt phosphorylates PGC-1 α at Ser 570. **a**, *In vitro* Akt kinase assay of indicated recombinant proteins. WT, wild type. **b**, H4IIE cells expressing PGC-1 α or S570A were treated with LY294002 (LY) or insulin as indicated. Lysates were blotted with anti-pPGC-1 α S570 or anti-PGC-1 α . **c**, H4IIE cells expressing PGC-1 α or GFP were treated as indicated. Lysates were subjected to immunoprecipitation (IP) and immunoblotting (IB). **d**, H4IIE cells were treated as indicated and lysates were immunoprecipitated with anti-pPGC-1 α S570 antibody followed by blotting with anti-PGC-1 α antibody. **e**, Primary hepatocytes from wild-type or *Akt2* $^{-/-}$ mice overexpressing PGC-1 α were treated with insulin with or without Akt

inhibitor (Akt1-1/2) as indicated. Lysates were probed with anti-pPGC-1 α S570 and with anti-PGC-1 α antibody. **f**, Liver nuclear extracts from wild-type or *Akt2* $^{-/-}$ mice were subjected to immunoprecipitation with anti-pPGC-1 α S570, followed by immunoblotting with anti-PGC-1 α antibody (upper panel). The lower panel shows the nuclear extracts blotted with an antibody recognizing total PGC-1 α . On the right is a quantification of the immunoblots, in which the immunoreactivity with the phospho-specific antibody has been normalized to the total PGC-1 α present in the extract. Values are expressed as means \pm s.e.m. ($n = 8$; asterisk, $P < 0.001$ compared with no insulin).

Another prediction of a model in which Akt directly inhibits PGC-1 α is that Akt should also be able to oppose actions of PGC-1 α on genes that are transcribed independently of FoxO1. In H4IIE cells, SHP mRNA levels decreased markedly on expression of MyrAkt, and were elevated about tenfold after overexpression of PGC-1 α or S570A (Fig. 2a). MyrAkt counteracted the effects of wild-type PGC-1 α but not S570A on the expression of SHP. PGC-1 α also coactivates estrogen-related receptor- α to promote the transcription of MCAD, thereby stimulating the β -oxidation of fatty acids^{19,20}. MCAD mRNA levels were increased about fourfold when PGC-1 α or S570A was overexpressed. Similarly, coexpression of MyrAkt antagonized the

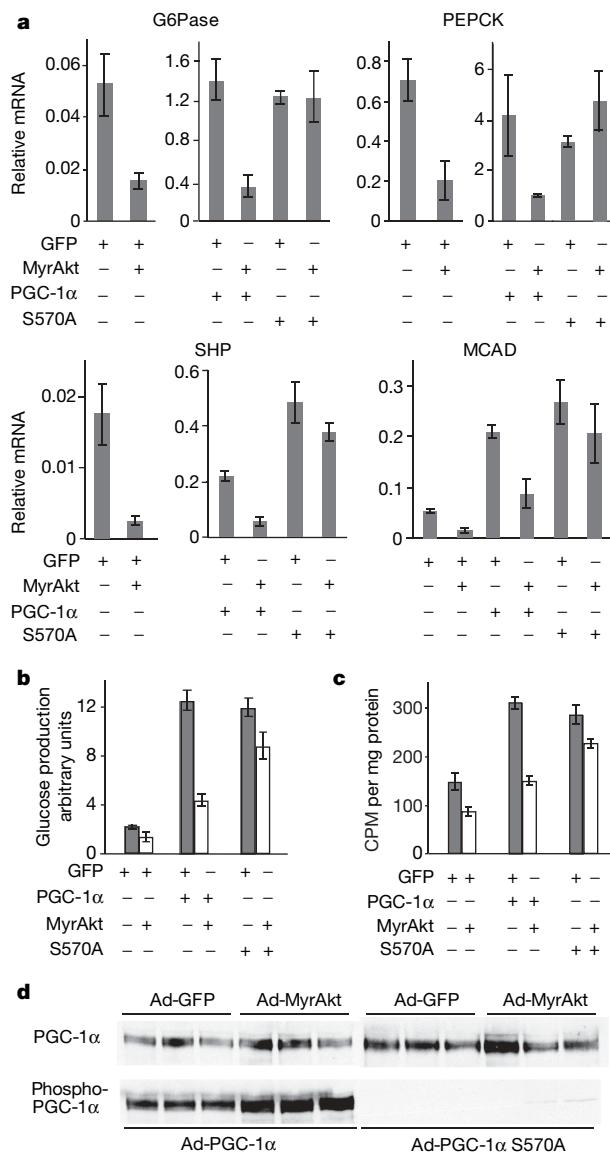


Figure 2 | Induction of gene expression, glucose production and β -oxidation by PGC-1 α S570A is resistant to inhibition by Akt. **a**, RNA from H4IIE cells infected with indicated adenoviruses was analysed for mRNA encoding G6Pase, PEPCK, SHP and MCAD by quantitative RT-PCR. Values are normalized to mRNA encoding glyceraldehyde-3-phosphate dehydrogenase and expressed as means \pm s.d. from two experiments performed in duplicate. **b**, Glucose production was assayed in primary hepatocytes infected with viruses as indicated. Values are normalized to the total cell protein and expressed as means \pm s.d. from one experiment performed in triplicate. **c**, β -oxidation of palmitate was assayed in primary hepatocytes infected with indicated viruses. Values are normalized to the total cell protein and expressed as means \pm s.d. from two experiments performed in triplicate. **d**, Immunoblots from representative experiments whose data are shown in **b** and **c**.

increase in MCAD induced by PGC-1 α but not that by S570A (Fig. 2a). MCAD is a key enzyme in the regulation of β -oxidation of fatty acids. To examine whether a change in the gene expression has an effect on the lipid metabolism, β -oxidation was monitored in primary hepatocytes. As shown in Fig. 2c, expression of PGC-1 α or S570A stimulated palmitate oxidation about twofold; coexpression of MyrAkt significantly blunted palmitate oxidation induced by PGC-1 α , but to a smaller extent by S570A. Taken together, these results indicate that Akt regulates both gluconeogenesis and the β -oxidation of fatty acids through the phosphorylation and inhibition of PGC-1 α .

Consistent with the data shown in Fig. 2a was our observation that levels of mRNA encoding G6Pase, PEPCK, MCAD and SHP were increased in the livers of mice overexpressing PGC-1 α or S570A, as delivered *in vivo* by adenovirus-mediated gene transfer (Fig. 3a). Coexpression of MyrAkt abrogated the effect of PGC-1 α but not that of S570A on these mRNAs. Whereas the induction of MCAD and SHP expression by S570A was totally resistant to inhibition by Akt, the Akt-dependent repression of G6Pase and PEPCK expression induced by S570A was incomplete. These results strongly suggest that the Akt/PGC-1 α pathway is operative in the intact animal. In concert with the changes in the gluconeogenic gene expression, infection of mice with adenovirus expressing PGC-1 α or S570A produced fasting

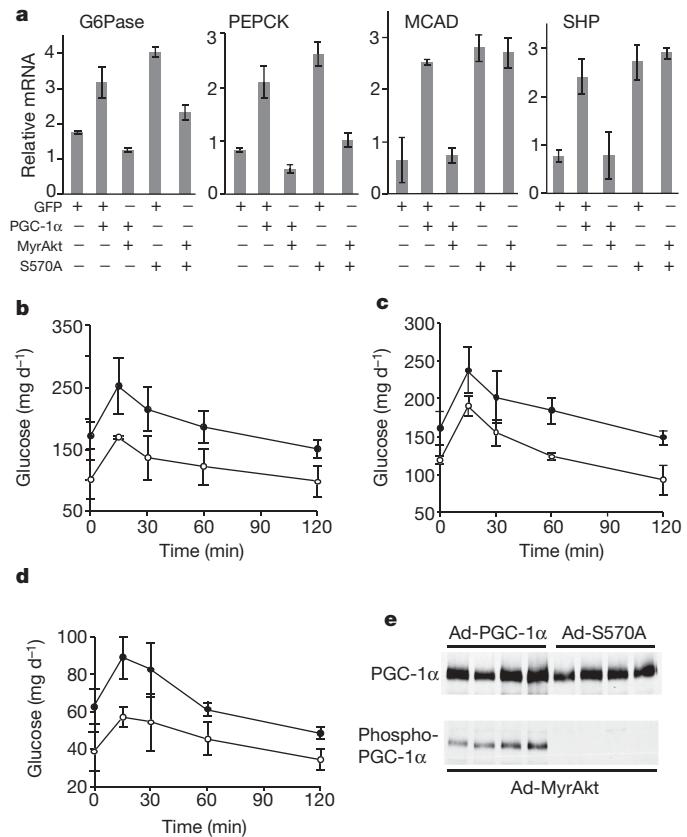


Figure 3 | Akt inhibits hepatic gene expression and glucose production in vivo by the phosphorylation of PGC-1 α at Ser 570. **a**, mRNA levels for G6Pase, PEPCK, MCAD and SHP were analysed by real-time RT-PCR in the livers of mice infected with the indicated viruses. Values represent means \pm s.d. from four animals in each group. **b**, A pyruvate tolerance test of mice expressing PGC-1 α (filled circles) or GFP (open circles). **c**, A pyruvate tolerance test of mice expressing S570A (filled circles) or GFP (open circles). Values in **b** and **c** represent means \pm s.d. from four animals in each group. **d**, A pyruvate tolerance test of mice coexpressing PGC-1 α plus MyrAkt (open circles) or S570A plus MyrAkt (filled circles). Values represent means \pm s.d. from two experiments with eight animals in each group. **e**, Western blots of liver homogenates from a representative experiment in which mice were infected with recombinant adenoviruses as indicated.

hyperglycaemia and elevated levels of blood glucose after the administration of pyruvate (Fig. 3b, c). Most importantly, blood glucose concentrations before and after pyruvate load were higher in mice coexpressing MyrAkt and S570A than in those expressing MyrAkt and PGC-1 α (Fig. 3d). Blood glucose concentrations were lower in mice expressing Akt in the liver than in controls expressing green fluorescent protein (GFP), whether or not PGC-1 α was also expressed, as reported previously²⁵. Taken together, these results show that phosphorylation of PGC-1 α at Ser 570 contributes significantly to the ability of Akt to control glucose homeostasis in mice.

PGC-1 α activates gene expression by recruitment to promoter regions through interaction with transcription factors. PGC-1 α was associated with the promoter region of *G6Pase*, an interaction that was ablated by coexpression of MyrAkt (Fig. 4a). Similarly, S570A also occupied the *G6Pase* promoter, but coexpression of MyrAkt could not inhibit its binding. If it is correct that Akt inhibits the activity of PGC-1 α through Ser 570 phosphorylation, this regulation should be applied to other promoters, in particular those independent of FoxO1. As shown in Fig. 4a, both PGC-1 α and S570A bound to the promoter region of *MCAD*. Coexpression of MyrAkt abrogated the binding of PGC-1 α but not that of S570A. These results show that the suppression of PGC-1 α activity by Akt is mediated by displacement of the former from chromatin, resulting in reduced transcriptional coactivation.

The prevailing model of the suppression of hepatic gluconeogenesis after a meal implies that insulin performs signalling through the Akt-mediated phosphorylation of FoxO1 to induce its export from the nucleus such that the transcription factor can no longer interact with PGC-1 α at target genes⁹. We show here that there is a secondary level of control of gluconeogenic genes by Akt that is mediated by its direct, inhibitory phosphorylation of PGC-1 α . This is consistent with a recent report showing that PGC-1 α coactivates HNF4 in addition to FoxO1 in the stimulation of *G6Pase* expression²⁶. Another consequence of the direct inhibition of PGC-1 α by Akt is a novel signalling pathway by which insulin suppresses fatty acid oxidation in the liver. The obvious implication of these data are that insulin can signal

through Akt to a network of genes not limited to those controlled by FoxO1 (Fig. 4c). Certainly, the inhibition of PGC-1 α and FoxO1 by Akt would work synergistically to curb the transcription of *PEPCK* and *G6Pase*, but phosphorylation of PGC-1 α also suppresses an entirely separate set of genes related to lipid catabolism. This is likely to be important to the hepatic response in the transition from fasting to the fed state, but it also raises the possibility that insulin might directly antagonize some of the effects of PGC-1 α in other tissues, such as skeletal muscle or brown adipose tissue.

METHODS SUMMARY

Immunoblotting and precipitation were performed as described previously²⁷. Glucose production and β -oxidation of fatty acids in primary hepatocytes was measured as described^{23,28,29}. For the preparation of liver nuclear extracts by sucrose ultracentrifugation, mice were fasted for 18 h, then injected intraperitoneally with insulin (2 U kg^{-1} body weight), and the liver was harvested 10 min later. H4IIE cells infected with the indicated adenovirus were subjected to chromatin immunoprecipitation analysis with a kit from Upstate, in accordance with the manufacturer's instructions. Chromatin was immunoprecipitated for 15 h with 6 μg of anti-PGC-1 α antibody (H-300; Santa Cruz) at 4°C . Purified DNA was analysed by real-time polymerase chain reaction with SYBR Green Mx3000 (Stratagene). Amplification of GCK promoter served as a negative control. The relative amount of a promoter DNA was calculated by the comparative C_t method. Values were normalized to TATA-binding protein DNA in the chromatin immunoprecipitates, with transthyretin as an invariant control. For *in vivo* expression of genes in liver, adenoviruses were injected through the jugular vein.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 January; accepted 16 April 2007.

Published online 6 June 2007.

- Bouche, C., Serdy, S., Kahn, C. R. & Goldfine, A. B. The cellular fate of glucose and its relevance in type 2 diabetes. *Endocr. Rev.* 25, 807–830 (2004).
- McGarry, J. D. Banting Lecture 2001: Dysregulation of fatty acid metabolism in the etiology of type 2 diabetes. *Diabetes* 51, 7–18 (2002).
- Zimmet, P., Alberti, K. G. & Shaw, J. Global and societal implications of the diabetes epidemic. *Nature* 414, 782–787 (2001).
- Diraison, F., Moulin, P. & Beylot, M. Contribution of hepatic de novo lipogenesis and reesterification of plasma non esterified fatty acids to plasma triglyceride synthesis during non-alcoholic fatty liver disease. *Diabetes Metab.* 29, 478–485 (2003).
- Donnelly, K. L. et al. Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *J. Clin. Invest.* 115, 1343–1351 (2005).
- Van Gaal, L. F., Mertens, I. L. & De Block, C. E. Mechanisms linking obesity with cardiovascular disease. *Nature* 444, 875–880 (2006).
- Eckel, R. H., Kahn, R., Robertson, R. M. & Rizza, R. A. Preventing cardiovascular disease and diabetes: a call to action from the American Diabetes Association and the American Heart Association. *Circulation* 113, 2943–2946 (2006).
- Lin, J., Handschin, C. & Spiegelman, B. M. Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab.* 1, 361–370 (2005).
- Puigserver, P. et al. Insulin-regulated hepatic gluconeogenesis through FOXO1–PGC-1 α interaction. *Nature* 423, 550–555 (2003).
- Barthel, A., Schmoll, D. & Untermaier, T. G. FoxO proteins in insulin action and metabolism. *Trends Endocrinol. Metab.* 16, 183–189 (2005).
- Liao, J., Barthel, A., Nakatani, K. & Roth, R. A. Activation of protein kinase B/Akt is sufficient to repress the glucocorticoid and cAMP induction of phosphoenolpyruvate carboxykinase gene. *J. Biol. Chem.* 273, 27320–27324 (1998).
- Cho, H. et al. Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB β). *Science* 292, 1728–1731 (2001).
- Kohn, A. D., Summers, S. A., Birnbaum, M. J. & Roth, R. A. Expression of a constitutively active Akt Ser/Thr kinase in 3T3-L1 adipocytes stimulates glucose uptake and glucose transporter 4 translocation. *J. Biol. Chem.* 271, 31372–31378 (1996).
- Ueki, K. et al. Potential role of protein kinase B in insulin-induced glucose transport, glycogen synthesis, and protein synthesis. *J. Biol. Chem.* 273, 5315–5322 (1998).
- Brazil, D. P., Yang, Z. Z. & Hemmings, B. A. Advances in protein kinase B signalling: AKTion on multiple fronts. *Trends Biochem. Sci.* 29, 233–242 (2004).
- Barthel, A. & Schmoll, D. Novel concepts in insulin regulation of hepatic gluconeogenesis. *Am. J. Physiol. Endocrinol. Metab.* 285, E685–E692 (2003).
- Nakae, J., Kitamura, T., Silver, D. L. & Accili, D. The forkhead transcription factor Foxo1 (Fkhr) confers insulin sensitivity onto glucose-6-phosphatase expression. *J. Clin. Invest.* 108, 1359–1367 (2001).

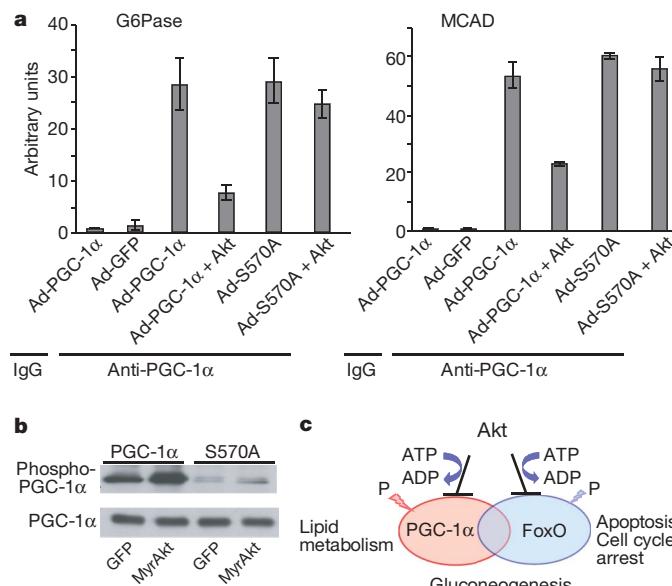


Figure 4 | Phosphorylation of PGC-1 α at Ser 570 is required for Akt to inhibit recruitment of PGC-1 α to chromatin. **a**, Lysates from H4IIE cells infected with indicated adenoviruses were subjected to chromatin immunoprecipitation analysis as described in Methods Summary. Values are expressed as means \pm s.d. from three independent experiments performed in duplicate. **b**, A representative immunoblot of the chromatin immunoprecipitation experiments. **c**, A model illustrating overlapping networks of control of gene expression by insulin-stimulated Akt phosphorylation of PGC-1 α and FoxO1.

18. Kanaya, E., Shiraki, T. & Jingami, H. The nuclear bile acid receptor FXR is activated by PGC-1 α in a ligand-dependent manner. *Biochem. J.* **382**, 913–921 (2004).
19. Schreiber, S. N., Knutti, D., Brogli, K., Uhlmann, T. & Kralli, A. The transcriptional coactivator PGC-1 regulates the expression and activity of the orphan nuclear receptor estrogen-related receptor α (ERR α). *J. Biol. Chem.* **278**, 9013–9018 (2003).
20. Huss, J. M., Kopp, R. P. & Kelly, D. P. Peroxisome proliferator-activated receptor coactivator-1 α (PGC-1 α) coactivates the cardiac-enriched nuclear receptors estrogen-related receptor- α and - γ . Identification of novel leucine-rich interaction motif within PGC-1 α . *J. Biol. Chem.* **277**, 40265–40274 (2002).
21. Pilegaard, H., Saltin, B. & Neufer, P. D. Exercise induces transient transcriptional activation of the PGC-1 α gene in human skeletal muscle. *J. Physiol. (Lond.)* **546**, 851–858 (2003).
22. Puigserver, P. & Spiegelman, B. M. Peroxisome proliferator-activated receptor- γ coactivator 1 α (PGC-1 α): transcriptional coactivator and metabolic regulator. *Endocr. Rev.* **24**, 78–90 (2003).
23. Yoon, J. C. et al. Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* **413**, 131–138 (2001).
24. Obata, T. et al. Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J. Biol. Chem.* **275**, 36108–36115 (2000).
25. Ono, H. et al. Hepatic Akt activation induces marked hypoglycemia, hepatomegaly, and hypertriglyceridemia with sterol regulatory element binding protein involvement. *Diabetes* **52**, 2905–2913 (2003).
26. Schilling, M. M., Oeser, J. K., Boustead, J. N., Flemming, B. P. & O'Brien, R. M. Gluconeogenesis: re-evaluating the FOXO1-PGC-1 α connection. *Nature* **443**, E10–E11 (2006).
27. Li, X., Yost, H. J., Virshup, D. M. & Seeling, J. M. Protein phosphatase 2A and its B56 regulatory subunit inhibit Wnt signaling in *Xenopus*. *EMBO J.* **20**, 4122–4131 (2001).
28. Finck, B. N. et al. Lipin 1 is an inducible amplifier of the hepatic PGC-1 α /PPAR α regulatory pathway. *Cell Metab.* **4**, 199–210 (2006).
29. Wolfrum, C., Besser, D., Luca, E. & Stoffel, M. Insulin regulates the activity of forkhead transcription factor Hnf-3 β /Foxa-2 by Akt-mediated phosphorylation and nuclear/cytosolic localization. *Proc. Natl. Acad. Sci. USA* **100**, 11624–11629 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Q. Chu for her help with the animal experiments, and P. Puigserver, M. Accili and B. Finck for providing reagents and advice. This work was supported by NIH grants to M.J.B. and X.L.

Author Contributions X.L. performed all the experiments, which were designed by X.L. and M.J.B. B.M. assisted in experiments with animals, and Q.G. generated the phospho-specific antiserum. The paper was written by X.L. and M.J.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.J.B. (birnbaum@mail.med.upenn.edu).

METHODS

Plasmids, adenoviruses, antibodies and viral infection. Expression plasmids for PGC-1 α , pGEX-GST-fusion PGC-1 α fragments, and PGC-1 α adenoviruses were a gift from B. Spiegelman. PGC-1 α mutants were generated with the Quickchange mutagenesis kit (Stratagene). Flag-tagged PGC-1 α and S570A adenoviruses were amplified by the University of Iowa Transgenics Core (Iowa City, IA, USA). Rabbit anti-PGC-1 α and anti-phospho-PGC-1 α were generated by Cell Signalling Technologies by using GST fused to PGC-1 α or a Ser 570 phosphorylated peptide, respectively.

Cell culture and treatment. All tissue culture media and antibiotics were purchased from Invitrogen. For insulin treatment, H4IIe cells were serum-starved for 16–18 h, followed by treatment with 100 nM insulin for 20 min. C57BL/6J mice were purchased from the Jackson Laboratory, and *Akt2* knockout mice were described previously¹². Primary hepatocytes were incubated with 1 μ M insulin, 200 μ M dexamethasone and 100 nM 3,3,5-triiodo-L-thyronine for 16–18 h, then exposed to recombinant adenovirus and incubated in insulin-free medium for a further 16–18 h. Cells were treated with 100 nM insulin for 20 min.

In vitro kinase assay. Kinase reaction (40 μ l) was reconstituted with 1 unit of constitutively active Akt1 (Upstate), 2 μ g of GST-fusion PGC-1 α fragment, [32 P]ATP (5 μ Ci), 10 mM ATP and kinase buffer (Cell Signalling Technologies). The reaction was incubated at 30 °C for 20 min and resolved by SDS-PAGE, followed by phosphorimager analysis.

Glucose production assay. Primary hepatocytes were infected with adenoviruses for 2 days before incubation for 3 h in 1 ml of glucose-free DMEM medium (pH 7.4), without phenol red, supplemented with 20 mM sodium lactate and 2 mM pyruvate²³. The medium was assayed for glucose with a colorimetric kit (Invitrogen). Values were normalized to the protein content of the whole-cell lysates.

Quantitative real-time PCR. Total RNA was prepared from H4IIe cells and from primary hepatocytes with the use of the RNeasy Mini Kit (Qiagen) and checked for its integrity by agarose-gel electrophoresis. cDNA was synthesized with random decamers with the RetroScript Kit (Ambion) and mixed with SYBR Green PCR Master Mix (Roche Diagnostics) and sets of gene-specific primers. The resulting mixture was subjected to real-time RT-PCR quantification with the Light Cycler System (Roche Diagnostics). A melting curve from 65 to 95 °C (0.05 °C s⁻¹) at the end of the reaction was used to check the purity and nature of the product. In all cases, a single PCR product with the predicted size was detected by agarose-gel electrophoresis. All reactions were performed in duplicate. The relative amounts of specific transcripts were calculated by using the comparative C_t method. Glyceraldehyde-3-phosphate dehydrogenase and trans-thyretin mRNAs were used as invariant controls.

Animal experiments. C57BL/6J mice (male, 23–25 g body weight) were purchased from the Jackson Laboratory and infected with adenoviruses by injection into the jugular vein. Adenoviruses encoding GFP or MyrAkt were given to mice at the dose of 2×10^9 virus particles g⁻¹, and Ad-PGC-1 α or S570A was administered at 2×10^8 plaque-forming units g⁻¹. Four days after virus injection, animals were fasted overnight and subjected to a pyruvate tolerance test. Pyruvate was injected intraperitoneally at a dose of 2 g kg⁻¹ and plasma glucose levels were measured from tail-vein blood with the Ascensia ELITE XL Blood Glucose Monitoring System (Bayer). Animals were then killed and livers were removed and snap-frozen. Total RNA was prepared from the frozen livers with Trizol reagent and the RNeasy Mini Kit. The resulting RNA samples were subjected to real-time RT-PCR analysis for expression of *G6Pase*, *PEPCK*, *MCAD* and *SHP*.

Palmitate oxidation. β -oxidation of fatty acids was measured as described²⁸. Primary murine hepatocytes infected with the indicated viruses were cultured in 0.5% DMEM without FBS for 16 h and labelled for 2 h with a [9,10-³H]palmitate/BSA mixture (3 μ Ci ml⁻¹) in the presence of 20 μ M unlabelled palmitate at 37 °C. After incubation, the medium was removed and proteins were precipitated with 10% trichloroacetic acid. The resulting supernatant was neutralized by NaOH and applied to ion-exchange columns (Dowex 1X8-200; Sigma) and the tritiated water was recovered. An aliquot of the eluates was then used for scintillation counting. Values were normalized to the total protein content of the whole-cell lysates.

LETTERS

Energy-dependent regulation of cell structure by AMP-activated protein kinase

Jun Hee Lee^{1,2*}, Hyongjong Koh^{1,2*}, Myungjin Kim^{1,2*}, Yongsung Kim^{1,2}, Soo Young Lee^{1,2}, Roger E. Karess³, Sang-Hee Lee⁴, Minho Shong⁵, Jin-Man Kim⁶, Jaeseob Kim^{2,7} & Jongkyeong Chung^{1,2}

AMP-activated protein kinase (AMPK, also known as SNF1A) has been primarily studied as a metabolic regulator that is activated in response to energy deprivation¹. Although there is relatively ample information on the biochemical characteristics of AMPK, not enough data exist on the *in vivo* function of the kinase. Here, using the *Drosophila* model system, we generated the first animal model with no AMPK activity and discovered physiological functions of the kinase. Surprisingly, AMPK-null mutants were lethal with severe abnormalities in cell polarity and mitosis, similar to those of *lkb1*-null mutants. Constitutive activation of AMPK restored many of the phenotypes of *lkb1*-null mutants, suggesting that AMPK mediates the polarity- and mitosis-controlling functions of the LKB1 serine/threonine kinase. Interestingly, the regulatory site of non-muscle myosin regulatory light chain (MRLC; also known as MLC2)^{2,3} was directly phosphorylated by AMPK. Moreover, the phosphomimetic mutant of MRLC³ rescued the AMPK-null defects in cell polarity and mitosis, suggesting MRLC is a critical downstream target of AMPK. Furthermore, the activation of AMPK by energy deprivation was sufficient to cause dramatic changes in cell shape, inducing complete polarization and brush border formation in the human LS174T cell line, through the phosphorylation of MRLC. Taken together, our results demonstrate that AMPK has highly conserved roles across metazoan species not only in the control of metabolism, but also in the regulation of cellular structures.

The catalytic subunit of *Drosophila melanogaster* AMPK is a single orthologue of its human and yeast counterparts^{1,4} (Supplementary Fig. 1a), and is activated by LKB1 on energy deprivation (Supplementary Fig. 1b). By imprecise excision of the EP-element (enhancer- and promoter-containing P-element) from the *AMPK^{G505}* line, we generated AMPK-null mutant lines, *AMPK^{D1}* and *AMPK^{D2}* (Fig. 1a), whose authenticity was confirmed by Southern blot, polymerase chain reaction with reverse transcription (RT-PCR) and immunoblot analyses (Fig. 1b–d). Interestingly, all AMPK-null mutant flies were lethal before the mid-pupal stage and failed to enter adulthood (Fig. 1e, f), even in the presence of sufficient nutrients. Although transgenic expression of wild-type AMPK (*AMPK^{WT}*) allowed AMPK-null mutants to successfully develop into adults (Fig. 1e), the expression of kinase-dead AMPK (*AMPK^{KR}*) failed to rescue the lethality (Fig. 1e), demonstrating that the phosphotransferase activity of AMPK is crucial for its function. In summary, we found AMPK to be essential for normal development of *Drosophila*.

Therefore, we further investigated the developmental role of AMPK by generating AMPK-null germ-line clone (AMPK-GLC)

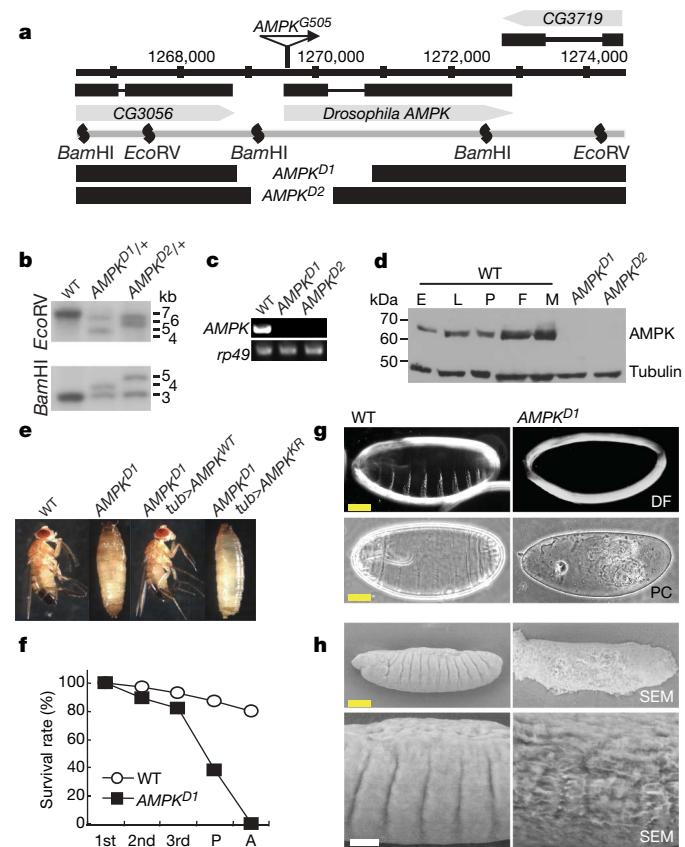


Figure 1 | AMPK is essential for normal development. **a**, Schematic genomic organization and restriction map of the AMPK locus. **b**, Southern blot analyses of AMPK in wild type (WT) and heterozygous AMPK mutants (*AMPK^{D1/+}*, *AMPK^{D2/+}*). **c**, RT-PCR analyses of AMPK in wild type and AMPK-null mutants (*AMPK^{D1}*, *AMPK^{D2}*). *rp49* was used as a loading control. **d**, Immunoblot analysis of AMPK in embryos (E), larvae (L), pupae (P), adult females (F) and adult males (M) of wild type and larvae of *AMPK^{D1}* and *AMPK^{D2}*. Tubulin was used as a loading control. **e**, Restored viability of AMPK-null (*AMPK^{D1}*) mutants by transgenic expression of wild-type (*AMPK^{WT}*) but not by kinase-dead (*AMPK^{KR}*) AMPK. *tub* (*tubulin*-*Gal4*+) was used as a wild-type control. See Supplementary Methods for details of nomenclature. **f**, Survival rates of wild type and AMPK-null mutants (*AMPK^{D1}*) during development: first, second and third instar larva, pupa (P), and adult (A). **g, h**, Cuticles (g) and surfaces (h) of wild-type and AMPK-GLC (*AMPK^{D1}*) embryos visualized by dark field (DF), phase contrast (PC), or scanning electron microscopy (SEM). Scale bars, yellow, 50 μ m; white, 20 μ m.

¹National Creative Research Initiatives Center for Cell Growth Regulation, and ²Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 373-1 Kusong-Dong Yusong-Gu, Taejon 305-701, Korea. ³Centre National de la Recherche Scientifique, Centre de Génétique Moléculaire, Avenue de la Terrasse, 91198 Gif sur Yvette, France. ⁴Division of Electron Microscopic Research, Korea Basic Science Institute, 52 Eoeun-Dong, Yusong-Gu, Taejon 305-333, Korea. ⁵Department of Internal Medicine, and ⁶Department of Pathology, Chungnam National University School of Medicine, 640 Daesa-Dong Chung-Gu, Taejon 301-721, Korea. ⁷GenExel, 373-1 Kusong-Dong, Yusong-Gu, Taejon 305-701, Korea.

*These authors contributed equally to this work.

embryos, which are completely deprived of both the maternal and zygotic AMPK proteins. Surprisingly, AMPK-GLC embryos never developed into larvae, showing the requirement of AMPK during embryogenesis. In AMPK-GLC embryos, cuticle structures were severely deformed, and ventral denticle belts were missing (Fig. 1g). Furthermore, the surface of AMPK-GLC embryos was roughened and the columnar structure of the epidermis was disorganized (Fig. 1h), implicating defects in underlying epithelial structures.

To examine the embryonic epithelial structures, we examined AMPK-GLC epithelia with various polarity markers. Bazooka (Baz, apical complex marker⁵) and β-catenin (Arm, adherens junction marker⁵) lost their apical localization and were found in various locations around the basolateral cell surfaces (Fig. 2a, c and Supplementary Fig. 2a). The Discs-large (Dlg, basolateral marker⁵) was also irregularly distributed throughout the epithelium in AMPK-GLC embryos (Fig. 2b, c). Moreover, actin staining demonstrated that the AMPK-GLC epithelium contained many unpolarized round cells that had lost contact with the underlying tissue (Supplementary Fig. 2b). This disorganization of epithelial structures was not a result of cell death, because it could not be restored by overexpression of apoptosis inhibitor p35 (ref. 6; Supplementary Fig. 3). In addition, wing discs of AMPK-null mutants also showed defective epithelial organization with ectopic actin structures in the basolateral region (Supplementary Fig. 4b, compared with 4a). These results indicate that AMPK is indispensable for epithelial integrity.

In addition, we found abnormally enlarged nuclei in some cells of AMPK-GLC embryos (Fig. 2d). Mitotic chromosome staining with anti-phospho-histone H3 (PH3) antibody demonstrated that

AMPK-GLC embryos frequently contained defective mitotic cells with lagging or polyploid chromosomes (Supplementary Fig. 5). Consistently, aceto-orcein staining of squashed AMPK-null larval brains revealed polypliody in ~30% of mitotic cells (Figs 2e and 3h), and anti-PH3 staining showed a highly increased amount of chromosome content in some of the neuroblasts (Fig. 3g, compared with 2j). These results indicate that AMPK is also required for the maintenance of genomic integrity.

Recently, it has been proposed that LKB1, a kinase upstream of AMPK (refs 1, 7), is involved in the regulation of epithelial polarity and mitotic cell division^{7–10}. Indeed, the abnormal polarity and mitosis phenotypes of *lkb1*-null mutants (Figs 2a–e, g, k and 3h, and Supplementary Fig. 4–7) were highly similar to those of AMPK-null mutants. To test whether AMPK mediates the polarity- and mitosis-controlling functions of LKB1, we expressed constitutively active AMPK (AMPK^{TD}), which is catalytically active even without phosphorylation by LKB1 (data not shown), in *lkb1*-null mutants. Remarkably, AMPK^{TD} suppressed the epithelial polarity defects (Fig. 2h and Supplementary Figs 4d and 7) and the genomic instability (Figs 2l and 3h) of *lkb1*-null mutants, suggesting that AMPK is a critical downstream mediator of LKB1, controlling mitosis and cell polarity.

To understand the molecular mechanism underlying the AMPK-dependent control of mitosis and cell polarity, we attempted to identify the downstream targets of AMPK. Intriguingly, MRLC, a critical molecule for the execution of mitosis and cell polarity establishment^{2,3,11,12}, contains a peptide sequence that can be phosphorylated by AMPK (ref. 13). Therefore, we performed various

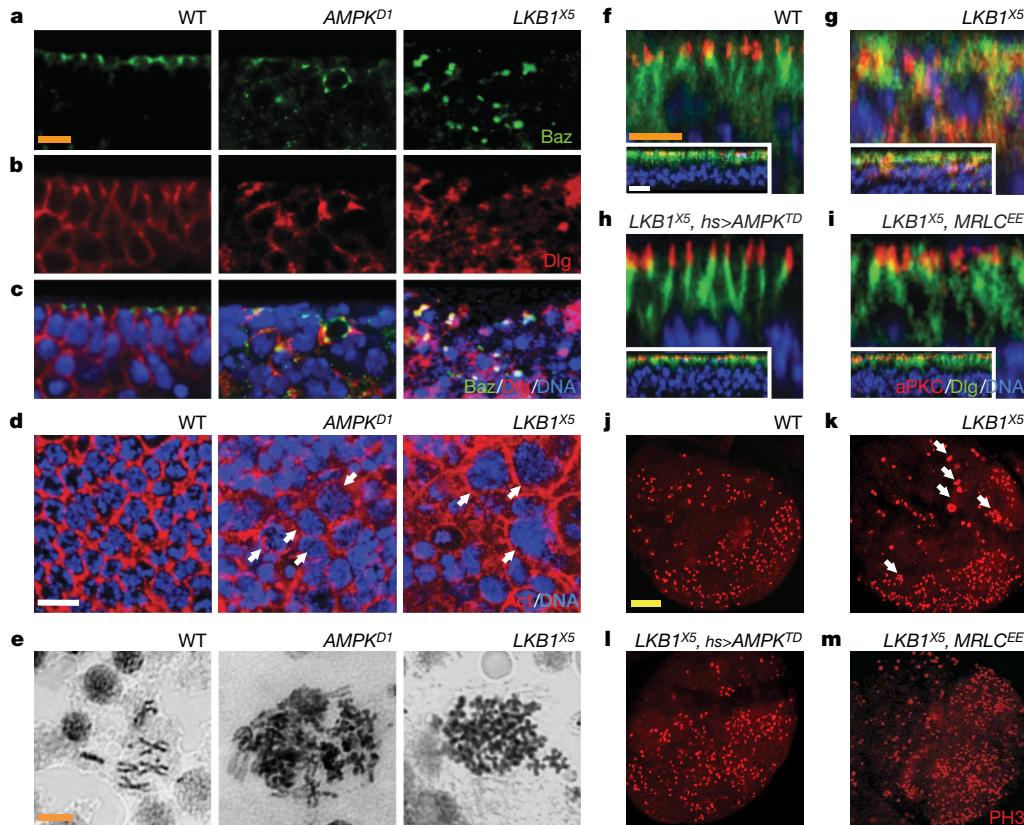


Figure 2 | AMPK and its upstream kinase LKB1 are required for epithelial and genomic integrity. **a–c**, Epithelia of wild-type, AMPK-GLC (*AMPK*^{D1}), and *lkb1*-GLC (*lkb1*^{X5}) embryos stained with anti-Baz (green) and anti-Dlg (red) antibodies and Hoechst 33258 (DNA, blue). **d**, Wild-type (WT), AMPK-GLC (*AMPK*^{D1}), and *lkb1*-GLC (*lkb1*^{X5}) embryos stained with Hoechst 33258 (DNA, blue) and anti-actin antibody (Act, red). Arrows indicate cells with an excessive DNA content. **e**, Aceto-orcein staining of wild-type, AMPK-null (*AMPK*^{D1}), and *lkb1*-null (*lkb1*^{X5}) larval neuroblasts.

f–m, Wing discs (**f–i**) and brain hemispheres (**j–m**) of wild-type, *lkb1*-null, *lkb1*-null expressing AMPK^{TD} (*lkb1*^{X5}, hs (heat shock)>*AMPK*^{TD}) and *lkb1*-null expressing *MRLC*^{EE} (*lkb1*^{X5}, *MRLC*^{EE}) larvae stained with anti-aPKC (apical complex marker⁵, green) and anti-Dlg (red) antibodies and Hoechst 33258 (DNA, blue) (**f–i**, or with anti-PH3 antibody (**j–m**). Vertical images were obtained by optical sectioning. Arrows indicate mitotic chromosomes with extreme polypliody. Scale bars, yellow, 50 μm; white, 10 μm; orange, 5 μm.

experiments to evaluate the ability of AMPK to phosphorylate MRLC. AMPK holoenzyme purified from rat liver strongly phosphorylated full-length MRLC, which was further enhanced by the addition of AMP (Fig. 3a, b). The phosphorylation of MRLC was more efficient than that of acetyl-CoA carboxylase 2 (ACC2), a representative substrate of AMPK (ref. 1) (Fig. 3c), indicating that MRLC is a good *in vitro* substrate of AMPK. We deduced that this phosphorylation is specifically performed by AMPK because Compound C, a specific inhibitor of AMPK, inhibited the phosphorylation, whereas ML-7, an inhibitor of another MRLC-phosphorylating kinase (MLCK)², did not (Fig. 3d). A mutant form of MRLC, whose regulatory phosphorylation site (corresponding to Thr 21/Ser 22 in *Drosophila*³ and Thr 18/Ser 19 in human²) was mutated into non-phosphorylatable alanines, was not phosphorylated by AMPK

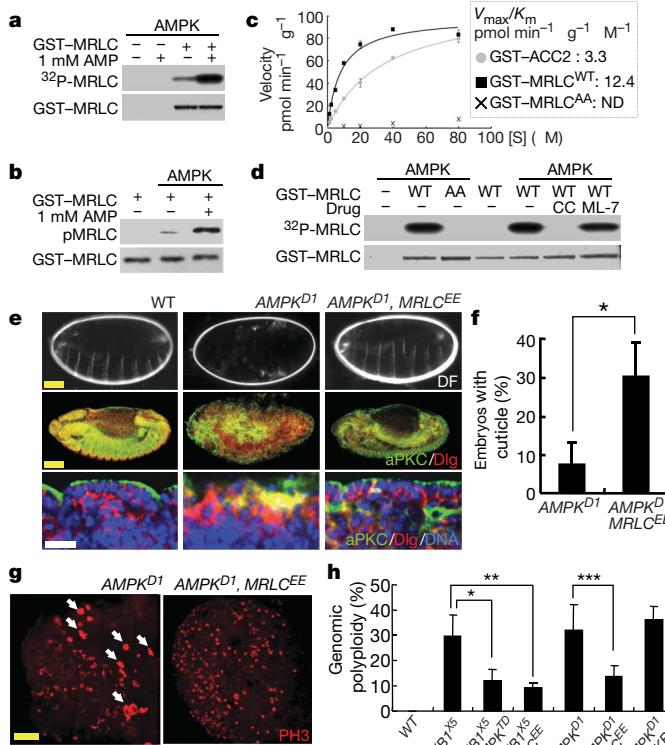


Figure 3 | AMPK phosphorylates MRLC to regulate mitosis and cell polarity. **a–d**, Phosphorylation of MRLC by AMPK. **a, b**, AMPK assay using human MRLC (GST–MRLC) substrate, with (+) or without (−) AMP. The phosphorylation was visualized by autoradiography (³²P–MRLC) or anti-phospho MRLC immunoblot (pMRLC), and the total MRLC by coomassie staining (a) or anti-GST immunoblot (b). **c**, Measurement of AMPK activity using the following substrates: human ACC2 (grey circle, GST–ACC2), wild-type (black square, GST–MRLC^{WT}) and T18A/S19A-mutant (black cross, GST–MRLC^{AA}) human MRLC. Error bars, s.d. of two experiments. [S], substrate concentration; ND, not determined. **d**, AMPK assay using wild-type and T18A/S19A-mutant (AA) GST–MRLC substrate. Compound C (CC) and ML-7 were added as indicated. Phosphorylated MRLC was visualized by autoradiography (³²P–MRLC), and total MRLC by coomassie staining. **e**, Cuticles and epithelial structures of wild-type, AMPK-GLC (AMPK^{D1}) and AMPK-GLC expressing MRLC^{EE} (AMPK^{D1}, MRLC^{EE}) embryos visualized by dark-field (DF) microscopy, anti-aPKC (green) and anti-Dlg (red) antibody staining, and Hoechst 33258 staining (DNA, blue). **f**, Quantification of cuticle phenotypes in AMPK-GLC (AMPK^{D1}) and AMPK-GLC expressing MRLC^{EE} (AMPK^{D1}, MRLC^{EE}) embryos (*P = 1.91 × 10⁻³). Error bars, s.d. of three experiments. **g**, The brain hemispheres from AMPK-null (AMPK^{D1}) and AMPK-null expressing MRLC^{EE} (AMPK^{D1}, MRLC^{EE}) larvae stained with anti-PH3 antibody. Arrows indicate mitotic chromosomes with extreme polypliody. **h**, Quantification of genomic polypliody in mitotic neuroblasts (*P = 1.36 × 10⁻²; **P = 5.32 × 10⁻³; ***P = 2.68 × 10⁻²). Error bars, s.d. of more than three experiments. Scale bars, yellow, 50 μm; white, 10 μm.

(Fig. 3c, d and Supplementary Fig. 8), suggesting that MRLC is exclusively phosphorylated at the regulatory phosphorylation site. Both the human and *Drosophila* forms of AMPK were able to phosphorylate MRLC from each of the respective species (Supplementary Figs 9, 10), which further demonstrates that the AMPK phosphorylation of MRLC is highly conserved between species.

Moreover, we found that MRLC phosphorylation is indeed regulated by AMPK *in vivo*. The phosphorylation of MRLC was dramatically reduced in AMPK- and *lkb1*-GLC epithelia when compared with the wild-type epithelia (Supplementary Fig. 6b, 11a, b), although the protein level of MRLC was unaffected (Supplementary Fig. 6c, 11e, f). The reduced phosphorylation of MRLC in the AMPK-GLC epithelia was completely restored by transgenic expression of AMPK but not by overexpression of LKB1 (Supplementary Fig. 11c, d). Furthermore, in *Drosophila* S2 cells, energy deprivation induced by 2-deoxyglucose (2DG) enhanced MRLC phosphorylation, which was suppressed by double-strand-RNA-mediated silencing of *lkb1* or AMPK (Supplementary Fig. 12). Collectively, these data strongly suggest that MRLC is specifically phosphorylated by AMPK both *in vitro* and *in vivo*.

To find out whether the phosphorylation of MRLC is critical for the physiological functions of AMPK, we expressed an active form of MRLC (ref. 3; MRLC^{EE}), whose regulatory phosphorylation site was mutated into phosphomimetic glutamates, in AMPK-GLC embryos. Strikingly, MRLC^{EE} rescued the epithelial polarity defects caused by the loss of AMPK (Fig. 3e), and increased the percentage of cuticle-forming embryos from ~10% to ~30% (Fig. 3f). MRLC^{EE} also restored the epithelial polarity defects of *lkb1*-null wing imaginal discs (Fig. 2i and Supplementary Fig. 7). Furthermore, the genomic polyploidy of AMPK- and *lkb1*-null larval brain neuroblasts was suppressed by the expression of MRLC^{EE} (Figs 2m and 3g, h). Therefore, we conclude that MRLC is a critical downstream target of AMPK controlling cell polarity and mitosis.

Notably, the larval brains of *MRLC* loss-of-function mutants (*spaghetti-squash*¹) showed extensive polyploidy (~40% of mitotic neuroblasts)¹⁴, and their imaginal discs showed severe disorganization in epithelial structure (Supplementary Fig. 13), similar to those of *lkb1*- and AMPK-null mutants. These phenotypic similarities further support our conclusion that MRLC is an important functional mediator of LKB1 and AMPK.

Finally, we questioned whether AMPK is critical for directing cell polarity in mammalian cells as well. To assess this, we examined whether the activation of AMPK by 2DG treatment (Supplementary Fig. 14) could induce polarization of unpolarized epithelial cells such as LS174T, which can be polarized by the activation of LKB1 (ref. 8), in a cell-autonomous manner. Surprisingly, on 2DG treatment, LS174T cells underwent a dramatic change in cell shape to have polarized actin cytoskeleton with a brush-border-like structure (Fig. 4a and Supplementary Figs 15 and 16a, f). Moreover, although brush border marker villin⁸ (Fig. 4b), apical marker CD66/CEA⁸ (Fig. 4c), and basal marker CD71/transferrin⁸ (Fig. 4d) were distributed throughout untreated cells, they became dramatically polarized on 2DG treatment (Fig. 4b–d), supporting that the activation of AMPK by energy deprivation is sufficient to induce complete polarization of LS174T cells.

We also found that the phosphorylation of MRLC by AMPK is involved in the energy-dependent polarization of LS174T cells. Phosphorylated MRLC was colocalized with the 2DG-induced polarized actin structures (Fig. 4e), and this phosphorylation, as well as the actin polarization, was suppressed by the AMPK-specific inhibitor Compound C (Fig. 4e and Supplementary Fig. 16a, f). Overexpression of dominant-negative AMPK (AMPK^{DN}) and short interfering (si)RNA-mediated inhibition of MRLC (siMRLC) also blocked the polarization (Fig. 4f, g and Supplementary Fig. 16e–g), although inhibition of Par-1, another downstream kinase of LKB1 (ref. 7), by Par-1 siRNA (siPar-1) or overexpression of dominant-negative Par-1 (Par-1^{DN}) failed to cause a block (Supplementary Fig. 16b–d, f, g). More strikingly, human MRLC^{EE} itself was sufficient to polarize

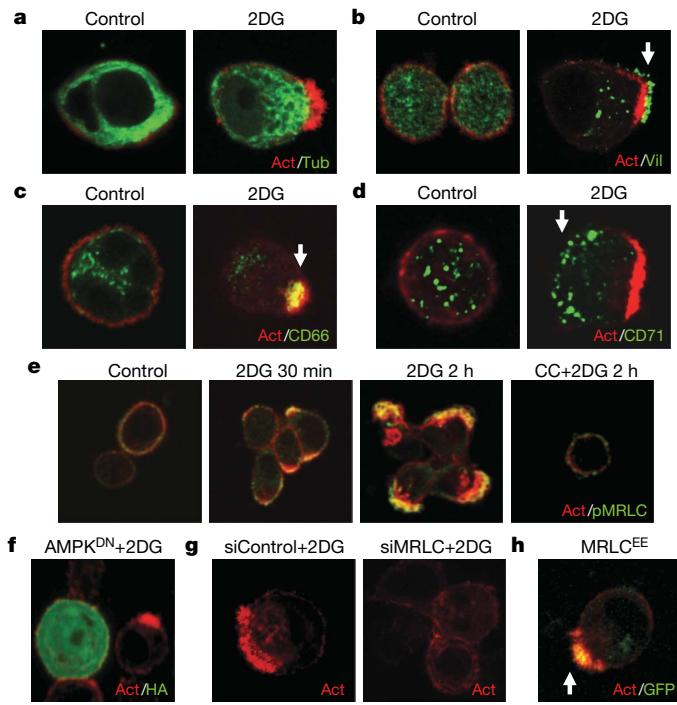


Figure 4 | Energy deprivation induces polarization of LS174T cells via MRLC phosphorylation by AMPK. **a–d**, Polarization of LS174T cells by energy deprivation. After a 2-l treatment with 2DG, LS174T cells were stained with anti-tubulin (Tub, green in **a**), anti-villin (Vil, green in **b**), anti-CD66/CEA (CD66, green in **c**), or anti-CD71/transferrin (CD71, green in **d**) antibodies, and TRITC-phalloidin (Act, red). Polarized markers are indicated by arrows. **e**, 2DG-induced phosphorylation of MRLC by AMPK. LS174T cells were stained with anti-phospho MRLC antibody (pMRLC, green) and TRITC-phalloidin (Act, red), after indicated treatments. **f**, Inhibition of the 2DG-induced polarization by human AMPK^{DN}. After a 2-h treatment of 2DG, cells were stained with TRITC-phalloidin (Act, red). AMPK^{DN} was detected by haemagglutinin (HA)-tag (green). **g**, Knockdown of MRLC suppressed the 2DG-induced polarization. After a 2-h treatment of 2DG, MRLC (siMRLC) or control (siControl), siRNA-treated cells were stained with TRITC-phalloidin (Act, red). **h**, Polarization of LS174T cells by expression of MRLC^{EE}. After 24 h of transfection, cells were stained with TRITC-phalloidin (Act, red). MRLC^{EE} detected by a green fluorescent protein (GFP)-tag (green) was co-localized with polarized actin structures (arrow). Original magnification, $\times 2,000$.

LS174T cells, even without energy deprivation (Fig. 4h and Supplementary Fig. 16h), showing that phosphorylation of MRLC is critical for the AMPK-dependent polarization.

Until now, the importance of AMPK has been limited to its role as a regulator of metabolism¹. However, by generating the first animal model with no AMPK activity, we characterized additional functions of AMPK: AMPK regulates mitotic cell division and epithelial polarity downstream of LKB1 by controlling the activity of MRLC through direct phosphorylation. Our findings revealed a link between energy status and cell structures, providing a new perspective to the diverse molecular function of AMPK. Further studies are needed on the cell-structure-controlling function of AMPK with respect to the various metabolic and physiological contexts, which may also help to understand AMPK-related diseases such as cancer and diabetes^{1,7,15}.

METHODS SUMMARY

Fly strains. AMPK^{D1} and AMPK^{D2} lines were generated by imprecise excision¹⁶ of AMPK^{G505} line (GenExel). UAS constructs were microinjected into *w¹¹¹⁸* embryos. *lkb1*- and MRLC-mutant lines were described previously^{3,14,17}. Other lines were obtained from the Bloomington Stock Center.

Microscopic analyses of *Drosophila* tissues. Dark field, phase contrast, and scanning electron microscopy (SEM) analyses were performed to visualize embryonic cuticle structure. Immunostainings with various antibodies were

performed to visualize embryonic epithelial structure. Brain squash preparation was used to analyse genomic integrity.

Kinase assays. Phosphotransferase activities were determined in a reaction mixture consisting of HEPES-Brij Buffer, ATP (γ^{32} P) and protein substrates. Protein levels were visualized by immunoblot or Coomassie staining. Phosphorylated proteins were visualized by phospho-specific immunoblot or 32 P-autoradiography. Incorporated phosphates were quantified using P81-filter-binding assay.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 January; accepted 10 April 2007.

Published online 7 May 2007.

- Kahn, B. B., Alquier, T., Carling, D. & Hardie, D. G. AMP-activated protein kinase: ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metabol.* **1**, 15–25 (2005).
- Matsumura, F. Regulation of myosin II during cytokinesis in higher eukaryotes. *Trends Cell Biol.* **15**, 371–377 (2005).
- Jordan, P. & Karess, R. E. Myosin light chain-activating phosphorylation sites are required for oogenesis in *Drosophila*. *J. Cell Biol.* **139**, 1805–1819 (1997).
- Pan, D. A. & Hardie, D. G. A homologue of AMP-activated protein kinase in *Drosophila melanogaster* is sensitive to AMP and is activated by ATP depletion. *Biochem. J.* **367**, 179–186 (2002).
- Knust, E. & Bossinger, O. Composition and formation of intercellular junctions in epithelial cells. *Science* **298**, 1955–1959 (2002).
- Hay, B. A., Wolff, T. & Rubin, G. M. Expression of baculovirus p35 prevents cell death in *Drosophila*. *Development* **120**, 2121–2129 (1994).
- Alessi, D. R., Sakamoto, K. & Bayascas, J. R. LKB1-dependent signalling pathways. *Annu. Rev. Biochem.* **75**, 137–163 (2006).
- Baas, A. F. et al. Complete polarization of single intestinal epithelial cells upon activation of LKB1 by STRAD. *Cell* **116**, 457–466 (2004).
- Martin, S. G. & St Johnston, D. A role for *Drosophila* LKB1 in anterior–posterior axis formation and epithelial polarity. *Nature* **421**, 379–384 (2003).
- Bettencourt-Dias, M. et al. Genome-wide survey of protein kinases required for cell cycle progression. *Nature* **432**, 980–987 (2004).
- Ivanov, A. I., Hunt, D., Utech, M., Nusrat, A. & Parkos, C. A. Differential roles for actin polymerization and a myosin II motor in assembly of the epithelial apical junctional complex. *Mol. Biol. Cell* **16**, 2636–2650 (2005).
- Edwards, K. A. & Kiehart, D. P. *Drosophila* nonmuscle myosin II has multiple essential roles in imaginal disc and egg chamber morphogenesis. *Development* **122**, 1499–1511 (1996).
- Michell, B. J. et al. Isoform-specific purification and substrate specificity of the 5'-AMP-activated protein kinase. *J. Biol. Chem.* **271**, 28445–28450 (1996).
- Karess, R. E. et al. The regulatory light chain of nonmuscle myosin is encoded by *spaghetti-squash*, a gene required for cytokinesis in *Drosophila*. *Cell* **65**, 1177–1189 (1991).
- Luo, Z., Saha, A. K., Xiang, X. & Ruderman, N. B. AMPK, the metabolic syndrome and cancer. *Trends Pharmacol. Sci.* **26**, 69–76 (2005).
- Lee, J. H. et al. *In vivo* p53 function is indispensable for DNA damage-induced apoptotic signaling in *Drosophila*. *FEBS Lett.* **550**, 5–10 (2003).
- Lee, J. H. et al. JNK pathway mediates apoptotic cell death induced by tumor suppressor LKB1 in *Drosophila*. *Cell Death Differ.* **13**, 110–1122 (2006).
- Lee, A. & Treisman, J. E. Excessive myosin activity in *Mbs* mutants causes photoreceptor movement out of the *Drosophila* eye disc epithelium. *Mol. Biol. Cell* **15**, 3285–3295 (2004).
- Sakurada, K. et al. Dynamics of myosin light chain phosphorylation at Ser19 and Thr18/Ser19 in smooth muscle cells in culture. *Am. J. Physiol.* **274**, 1563–1572 (1998).
- Park, J. et al. Mitochondrial dysfunction in *Drosophila* PINK1 mutants is complemented by parkin. *Nature* **441**, 1157–1161 (2006).
- Monfar, M. et al. Activation of pp70/85 S6 kinases in interleukin-2-responsive lymphoid cells is mediated by phosphatidylinositol 3-kinase and inhibited by cyclic AMP. *Mol. Cell. Biol.* **15**, 326–337 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank A. Wodarz, K. L. Guan, R. Fehon, C. Sunkel, R. Saint, H. Piwnica-Worms, D. St Johnston, C. Q. Doe, R. T. Moon, M. Montminy, L. Alphey and H. Clever, as well as the Bloomington Stock Center, Developmental Studies Hybridoma Bank and *Drosophila* Genomics Research Center, for kindly providing materials. We also thank the Korea Basic Science Institute for electron microscopy analyses and the Korean Cell Line Bank for LS174T cell stock. This research was supported by a National Creative Research Initiatives grant from the Korean Ministry of Science and Technology/KOSEF.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.C. (jchung@kaist.ac.kr).

METHODS

Drosophila strains. The AMPK^{G505} fly line with an EP-element inserted at the AMPK locus was obtained from the Genisis Collection of EP lines (GenExel), and subjected to standard imprecise excision^{16,17}. HA-tagged AMPK (AMPK^{WT}) and its mutant forms (AMPK^{KR} , AMPK^{TA} and AMPK^{TD}) were subcloned into the pUAST vector and microinjected into w^{1118} embryos. *lkb1*-null mutant (*lkb1*^{X5}), *UAS-lkb1*, *spaghetti-squash*¹ (*sqh*), and *MRLC*^{EE} (also referred to as *sqh*^{EE}) lines were described previously^{3,14,17}. The *Gal4* lines, balancer lines, *UAS-p35* line, and germ-line clone lines (FRT/FLP lines and dominant female-sterile lines) were obtained from the Bloomington Stock Center.

Antibodies. His-tagged full-length *Drosophila* AMPK protein was purified by nickel affinity-column and injected into rabbits to generate anti-*Drosophila* AMPK antibody. In addition, anti-*Drosophila* LKB1¹⁷, anti-phospho-Ser19 human MRLC (Cell Signaling, no. 3671 was used to detect phospho-Ser22 *Drosophila* MRLC¹⁸, and no. 3675 was used to detect phospho-Ser 19 human MRLC¹⁹), anti-full-length human MRLC (FL-172, Santa Cruz; this antibody also recognizes *Drosophila* MRLC), anti-phospho-Thr 172 human AMPK (40H9, Cell Signaling; this antibody also recognizes phospho-Thr 184 *Drosophila* AMPK⁴), anti-human AMPK (Cell Signaling), anti-human LKB1 (Upstate), anti-villin (CWVB1, Abcam), anti-CD66 (1C3, Abcam), fluorescein isothiocyanate (FITC)-conjugated anti-CD71 (L01.1, BD Biosciences), anti-HA (3F10 (Roche) was used for immunoblot, 12CA5 (Roche) was used for immunoprecipitation, and 6E2 (Cell Signaling) was used for immunocytochemistry.), anti-atypical PKC (Santa Cruz), anti-phospho-Ser10 histone H3 (Upstate), anti-Bazooka (a gift from A. Wodarz), anti-*Drosophila* Par-1 (a gift from D. St Johnston), anti-GST (Upstate), anti-human Par-1 (183.A10.A3, Upstate), anti-Scribble (a gift from C. Q. Doe), anti-Armadillo (N2 7A1, DSHB), anti-Discs large (4F3, DSHB), anti-Crumps (Cq5, DSHB), anti-*Drosophila* E-cadherin (DCAD2, DSHB), anti-Na⁺/K⁺ ATPase (a5, DSHB), anti-β-tubulin (E7, DSHB), and anti-actin (JLA20, DSHB) antibodies were used for immunostaining and immunoblot analyses. TRITC-labelled phalloidin (Sigma) and Hoechst 33258 (Sigma) were also used to visualize filamentous actins and DNA, respectively.

Confirmation of AMPK-null mutants. Deletion sites of AMPK^{D1} and AMPK^{D2} were determined by genomic PCR analyses. EP-element insertion in AMPK^{G505} mutant (triangle) and genomic deletions in AMPK^{D1} (1,268,785–1,270,743th basepair) and AMPK^{D2} (1,269,080–1,270,246th basepair) mutants are indicated in Fig. 1a. Nucleotide numbering in Fig. 1a was done according to the *Drosophila melanogaster* chromosome X sequence (release v5.1). For Southern blot analyses^{16,20}, 10 µg of total genomic DNA from w^{1118} , $\text{AMPK}^{D1}/+$, $\text{AMPK}^{D2}/+$ flies was digested with *EcoRV* or *BamHI* (KOSCHEM). *Drosophila* AMPK full-length ORF was used as a probe. For RT-PCR analyses^{16,17,20}, total RNA from the third instar larvae was extracted with Easy-BlueTM (Intron), and reversely transcribed using *Maxime* RT premix kit (Intron). Then, the following primers were used to amplify the AMPK transcripts by PCR: 5'-GATCACACGCGTCAAGGTGGCC-3' and 5'-GGTCTCGATGCACGATCATGTGCC-3'. *rp49* transcripts were amplified as previously described^{16,20}.

Microscopic examination of *Drosophila* embryos. Dark field and phase contrast analyses were performed using a light microscope (DM-R, Leica). Embryonic cuticles were prepared as previously described¹⁷. To quantify the number of embryos with cuticle for Fig. 3f, we performed three replicate experiments (60–90 embryos were observed for each experiment) to calculate the proportion of the number of embryos with ventral denticle belts to the total number of embryos. The average ratio of the three experiments is presented in per cent, as a bar graph, and the standard deviation is indicated as error bars. A *P*-value was calculated using one-way ANOVA analysis. For SEM analyses, embryos were dechorionated with 50% bleach, fixed with a mixture of 4% paraformaldehyde solution and heptane, then devitellinized with a methanol-heptane mixture. Rehydrated embryos were dried in air for 3 min and frozen for SEM analyses. SEM images were obtained by LEO 1455VP Electron Microscopy System in the VPSE (Variable Pressure Secondary Electron) mode.

Visualization of mitotic chromosomes in larval brain. Larval brains were dissected in isotonic saline. The brains were first incubated for 90 min in 5×10^{-5} M colchicine in saline, then were hypotonically shocked for 10 min in 0.5% sodium citrate. Finally, the brains were fixed and stained as described previously¹⁴. Cytological examination was performed under a light microscope (DM-R, Leica). For quantification of genomic polyploidy in Fig. 3h, we calculated the proportion of the number of polyploid mitotic neuroblasts to the total number of mitotic neuroblasts. We examined 400–700 mitotic cells from more than three different brains of the same genotype. The average of the ratios among brains is presented in per cent, as bar graphs, and the standard deviation is indicated as error bars. *P*-values were calculated using one-way ANOVA analyses.

In vitro kinase assay. The protein kinase assay was performed in a solution consisting of HEPES-Brij Buffer, 0.2 mM ATP (with $0.5 \mu\text{Ci } \mu\text{l}^{-1} \gamma^{32}\text{P}$ -ATP for radioactive assay), and 1 µg, or indicated amount, of protein substrate at

30 °C for 20 min. For Fig. 3d and Supplementary Fig. 10b, c, 10 µM Compound C or 10 µM ML-7 were added to the reaction mixture as indicated. For Fig. 3a, b, 0.3 mM AMP was added to the mixture as indicated. Except for Fig. 3a, b, all reaction mixtures contained 0.3 mM AMP. Except for Fig. 3c and Supplementary Fig. 8a, the assay samples were subjected to SDS-PAGE, and then autoradiography or immunoblotting was performed. For Fig. 3c and Supplementary Fig. 8a, the radioactive assay samples were subjected to P81-filter-binding assay²¹. V_{max} and K_m values were obtained by nonlinear regression analysis (curve fitting) using GraphPad Prism version 4.0 (Graphpad Software). V_{max} values (\pm s.e.m.) for GST-ACC2 and GST-MRLC^{WT} were $113.9 \pm 3.216 \text{ pmol } \mu\text{g}^{-1} \text{ min}^{-1}$ and $98.93 \pm 5.264 \text{ pmol } \mu\text{g}^{-1} \text{ min}^{-1}$, respectively. K_m values (\pm s.e.m.) of GST-ACC2 and GST-MRLC^{WT} were $34.37 \pm 2.114 \mu\text{M}$ and $8.001 \pm 1.417 \mu\text{M}$, respectively. Because kinetic parameters for GST-MRLC^{AA} did not converge, curve fitting was not possible.

Detailed methods on molecular biology, fly genetics, immunoblot, immunostaining, preparation of kinases and their substrates, *Drosophila* S2 cell culture, and mammalian cell culture are described in Supplementary Methods.

LETTERS

Mechanism of coupled folding and binding of an intrinsically disordered protein

Kenji Sugase^{1,2}, H. Jane Dyson¹ & Peter E. Wright¹

Protein folding and binding are analogous processes, in which the protein ‘searches’ for favourable intramolecular or intermolecular interactions on a funnelled energy landscape^{1,2}. Many eukaryotic proteins are disordered under physiological conditions, and fold into ordered structures only on binding to their cellular targets^{3–6}. The mechanism by which folding is coupled to binding is poorly understood, but it has been hypothesized on theoretical grounds that the binding kinetics may be enhanced by a ‘fly-casting’ effect, where the disordered protein binds weakly and non-specifically to its target and folds as it approaches the cognate binding site⁷. Here we show, using NMR titrations and ¹⁵N relaxation dispersion, that the phosphorylated kinase inducible activation domain (pKID) of the transcription factor CREB forms an ensemble of transient encounter complexes on binding to the KIX domain of the CREB binding protein. The encounter complexes are stabilized primarily by non-specific hydrophobic contacts, and evolve by way of an intermediate to the fully bound state without dissociation from KIX. The carboxy-terminal helix of pKID is only partially folded in the intermediate, and becomes stabilized by intermolecular interactions formed in the final bound state. Future applications of our method will provide new understanding of the molecular mechanisms by which intrinsically disordered proteins perform their diverse biological functions.

Following phosphorylation of Ser 133 in KID, CREB stimulates gene expression by association with the general transcriptional coactivator CBP (CREB binding protein) or its parologue p300 (refs 8, 9). CBP and p300 bind directly to pKID via the KIX domain¹⁰. Free pKID is intrinsically disordered but folds to form two α -helices, designated α_A and α_B , on binding to KIX (ref. 11). Phosphorylation is required for high-affinity binding, as are hydrophobic contacts between α_B and a shallow groove on KIX (refs 12–14). To elucidate the mechanism by which pKID folds on binding to KIX, we used ¹⁵N relaxation dispersion data^{15,16} measured as a function of pKID:KIX concentration ratio and ¹H-¹⁵N single quantum correlation (HSQC) titrations. These experiments identify weakly populated intermediate states, and provide insights into their structure, their population, and the kinetics of the coupled binding and folding processes.

The ¹H-¹⁵N HSQC spectrum of free pKID is characteristic of an unstructured protein¹¹. On titration of KIX into ¹⁵N-labelled pKID (Fig. 1), the intensity of the free pKID cross-peaks diminishes, and new resonances appear and grow in intensity at the chemical shifts of the pKID:KIX complex. Exchange between the free and bound states is slow on the chemical shift timescale, and cross-peaks corresponding to both states are observed at intermediate pKID:KIX concentration ratios. However, several free pKID cross-peaks shift in fast exchange at stoichiometric concentrations of KIX (Fig. 1, top two insets). Many of the affected resonances do not move directly towards

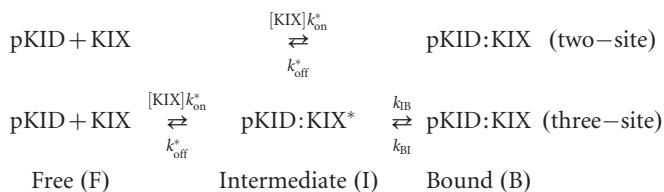
the bound cross-peak, which means that the shifts must arise from a secondary binding process to form a weak complex in fast exchange with free pKID. HSQC titrations using ¹⁵N-labelled KIX confirm that pKID binds in slow exchange to the primary high-affinity site and in fast exchange to additional non-specific hydrophobic sites on the surface of KIX, including the binding site for the transactivation domain of the mixed myeloid leukaemia protein (MLL) (Supplementary Figs S2, S3).

Evidence that the fast-exchange shifts of pKID resonances result from transient interactions with KIX comes from competition and mutagenesis experiments. Titration with the MLL activation domain, which forms a ternary complex with pKID and KIX (refs 17, 18), reverses the fast-exchange shifts in the pKID spectrum by blocking non-specific interactions that overlap the MLL binding site (Fig. 1, middle two insets). Some interactions appear to be native-like: substitution of Tyr 658 of KIX by Phe, which disrupts a critical intermolecular hydrogen bond to the phosphoserine in the high-affinity complex¹¹, specifically decreases the shifts for pSer 133 and Tyr 134 while leaving other pKID cross-peaks unaffected (Fig. 1, bottom two insets). Thus, in addition to binding at the high-affinity site, pKID explores an ensemble of weak complexes at multiple sites on the KIX surface. This behaviour is suggestive of a diffusional encounter complex, a transient state in which the two molecules are held together by fluctuating short-range interactions as they search for the mutually favourable orientations and contacts that stabilize the fully bound state^{19,20}.

In contrast to encounter complexes formed between globular proteins^{21,22}, where association is driven by complementary electrostatic interactions, the transient complexes between pKID and KIX appear to be stabilized predominantly by hydrophobic contacts. The pKID residues that experience the largest chemical shift changes are, with the exception of pSer 133, bulky hydrophobic residues (Tyr 134, Ile 137 and Leu 138) that lie on one face of helix α_B and are essential for high-affinity binding^{11,14} (Figs 1, 2). The magnitude and pattern of the shifts suggests that the α_B region is partly helical (up to 30%) in the low-affinity complex. Resonances from α_A are shifted only slightly by the fast exchange process.

To obtain further insights into the mechanism of coupled binding and folding, we performed ¹⁵N R_2 dispersion experiments, where R_2 is the transverse relaxation rate. The effective R_2 relaxation rates (R_2^{eff}) depend on the KIX concentration, showing that the observed dispersion is associated with the binding process (Fig. 3a). Dispersion cannot arise from folding transitions in unbound pKID, to form a minor population of a binding-competent conformer, because no relaxation dispersion is observed for free pKID (Supplementary Fig. S4). The dispersion curves at two spectrometer frequencies and at four pKID:KIX concentration ratios were fitted simultaneously with two- and three-site exchange models:

¹Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ²Suntory Institute for Bioorganic Research, 1-1 Wakayamada, Shimamoto-cho, Mishima-gun, Osaka 618-8503, Japan.



The two-site exchange model resulted in physically unrealistic parameters; the effective microscopic dissociation constant K_D^* derived from the fits (41 mM for pSer 133) is inconsistent with the macroscopic K_D (1.3 μM) determined by isothermal titration calorimetry²³. In contrast, the three-site exchange model fits better (reduced $\chi^2 = 1.12$ versus 1.60) and gives a mean K_D^* ($1.5 \pm 1.0 \mu\text{M}$) that is in excellent agreement with isothermal titration calorimetry (Table 1).

The dispersion curves for residues in the helical regions of bound pKID were analysed in clusters, with the folding and unfolding rate constants k_{IB} and k_{BI} fixed for all residues in each cluster. Three clusters were found (Table 1); residues 124–128 in α_A form a single cluster (I), while there are two clusters in α_B (clusters II and III, comprising residues 133–139 and 140–144, respectively) which appear to interact independently with KIX in both the encounter complex and the intermediate.

We simulated ^{15}N NMR spectra²⁴ using the parameters of Table 1. Even though the rate of exchange between bound and intermediate states ($k_{\text{ex,IB}}$) is fast ($k_{\text{ex,IB}} > \Delta\omega_{\text{IB}}$, where $\Delta\omega_{\text{IB}}$ is the chemical shift difference between the intermediate and bound states), the overall exchange process between free and fully bound states is slow (Fig. 3b),

as observed experimentally. Because exchange between I and B is fast, the ‘bound’ shift (given by $\Delta\omega^*$) is a population weighted average of the chemical shifts in the I and B states. Comparison of the chemical shift difference between the free and intermediate states, $\Delta\omega_{\text{FI}}$, to $\Delta\omega^*$ provides a measure of the extent of folding of each residue in the intermediate. For residues in cluster I, the average value of $\Delta\omega_{\text{FI}}/\Delta\omega^*$ is greater than 0.9, suggesting that the α_A helix is nearly fully folded. Although α_A residues (and Arg 131) make nearly native interactions in the intermediate, they experience additional millisecond timescale fluctuations that contribute to relaxation dispersion. In contrast, the average $\Delta\omega_{\text{FI}}$ for residues 133–138 and 141 in the α_B helix is only about 70% of the fully bound shift, showing that this helix is incompletely folded (Fig. 3c). Thus, in the intermediate, the critical hydrophobic residues that anchor the pKID α_B helix to the KIX hydrophobic groove continue to search for the favourable intermolecular contacts that stabilize the final bound state.

To elucidate the mechanism of folding of α_B , we looked for correlations between the chemical shifts determined from the dispersion data and equilibrium chemical shift changes measured from HSQC spectra, $\Delta\delta$ (Fig. 3d). A linear correlation is observed between $\Delta\omega^*$ and $\Delta\delta$. Importantly, $\Delta\omega^*$ correlates significantly better with the differences in chemical shift between the bound state and the ensemble of weak non-specific complexes, rather than with shift differences relative to free pKID. This strongly suggests that the transient encounter complexes are essential for productive binding, and evolve to the high-affinity complex without dissociation of pKID from KIX.

As α_B is incompletely folded in the intermediate identified from the dispersion experiments, we conclude that this state is a pKID folding

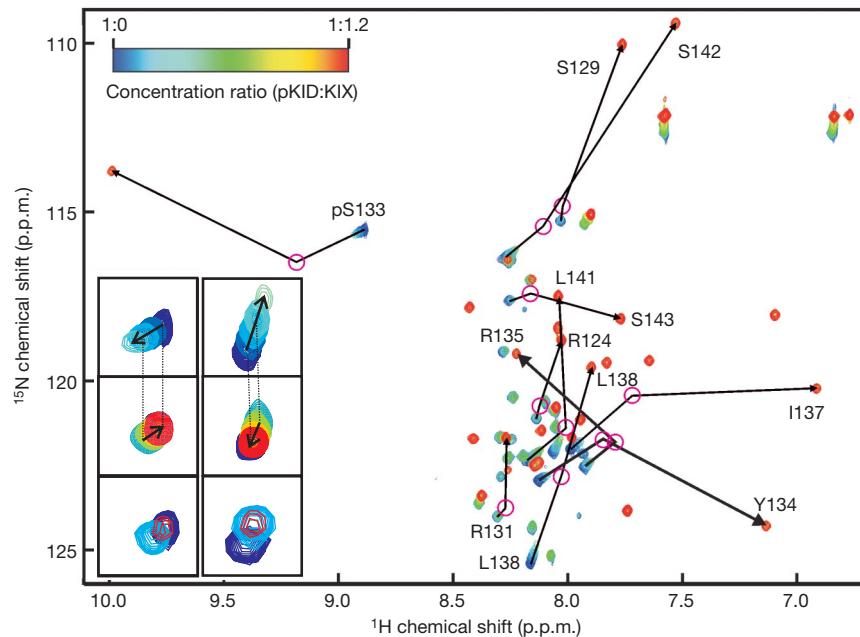


Figure 1 | Chemical shift changes on pKID binding to KIX. ^1H - ^{15}N HSQC titration of $[^{15}\text{N}]\text{-pKID}$ with unlabelled KIX in which pKID:KIX concentration ratios ranged from 1:0 to 1:1.2. The cross-peak colour changes gradually from blue (free) to red (bound) according to the concentration ratio. Exchange between the free and fully bound states of pKID is slow on the chemical shift timescale. At substoichiometric concentrations of KIX, several of the pKID cross-peaks shift away from their free positions in fast exchange. The open circles (magenta) represent the fully bound chemical shifts for the fast exchange process. These were estimated using the dissociation constant (680 μM) determined by fitting the shifts observed for KIX resonances in the presence of excess pKID (Supplementary Fig. S2). Selected resonances are labelled. An expanded view of the central part of this figure is shown in Supplementary Fig. S7. Insets show expanded views of the

cross-peaks for pSer 133 and Leu 138 at substoichiometric pKID:KIX ratios, as follows. Top row, fast-exchange shift of the resonances of pSer 133 (left) and Leu 138 (right) at pKID:KIX mole ratios of 1:0, 1:0.1, 1:0.2, 1:0.3, 1:0.4. The contours are coloured according to the scale at the top of the figure. Middle row, reversal of fast-exchange shifts for pSer 133 (left) and Leu 138 (right) on titration with MLL. The spectra correspond to pKID:KIX:MLL mole ratios of 1:0:0 (dark blue contours), 1:0:3:0 (cyan), 1:0:3:0.1 (green), 1:0:3:0.2 (yellow) and 1:0:3:0.3 (red). Bottom row, HSQC cross-peaks of pSer 133 (left) and Leu 138 (right) in the presence of wild-type KIX (cyan contours) and the Tyr658Phe mutant (red) at a pKID:KIX mole ratio of 1:0.1. For reference, the cross-peaks for free pKID are shown in dark blue. The scale is slightly expanded relative to the other insets.

intermediate that differs significantly from the encounter complex. Strictly, the coupled folding and binding process should be modelled as four-site exchange (Free \leftrightarrow Encounter \leftrightarrow Intermediate \leftrightarrow Bound). However, exchange of pKID between the free state and the encounter complex is fast and does not contribute to ^{15}N R_2 relaxation; dispersion arises from exchange between the encounter complex, the folding intermediate, and the fully bound state. At steady state, the steps F \leftrightarrow E \leftrightarrow I (where E designates the encounter complex) can be replaced by a single-step process with effective association and dissociation rate constants $k_{\text{on}}^* = k_{\text{on}} k_{\text{EI}} / (k_{\text{off}} + k_{\text{EI}})$ and $k_{\text{off}}^* = k_{\text{off}} k_{\text{IE}} / (k_{\text{off}} + k_{\text{EI}})$ (ref. 25).

The values of k_{on}^* show that all residues of pKID except Arg 131 bind to KIX at similar rates (average $k_{\text{on}}^* = 6.3 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$) to form the intermediate. The larger k_{on}^* for Arg 131 suggests a possible role for this residue in electrostatic steering²⁶. In contrast, the effective rate of dissociation from the intermediate, k_{off}^* , differs significantly for individual residues. Residues in α_A (cluster I; average $k_{\text{off}}^* = 12.7 \text{ s}^{-1}$) dissociate most slowly, whereas those in α_B have more than fivefold larger k_{off}^* values suggesting that they interact more weakly

with KIX in the intermediate. The intermolecular contacts made by α_B residues are stabilized following evolution of the intermediate to the final state. For the fully bound state, the effective microscopic dissociation constants K_D^* and the apparent dissociation rates k_{off} are the same for the residues in each cluster. Thus, the critical interfacial hydrophobic residues in the α_A and α_B helices dissociate from the fully bound state at the same rates (average k_{off} is 6.5 s^{-1} for Leu 128, Tyr 134, Leu 138 and Leu 141). The increased apparent k_{off} rates for Asp 140 and Asp 144 cannot represent physical dissociation of pKID from KIX because neighbouring residues have much smaller k_{off} . Rather, the dispersion data probably contain contributions from additional exchange processes, reflecting either transient local distortions of this region of the α_B helix or fluctuating electrostatic interactions between the Asp side chains and His 602 and Lys 606 of KIX.

Our studies provide insights into the mechanism by which an intrinsically disordered protein folds on binding to its target. Application of our method to additional complexes will be needed to establish whether coupled folding and binding is ubiquitous, or if alternative mechanisms, such as conformational selection (Supplementary Fig. S1), also operate.

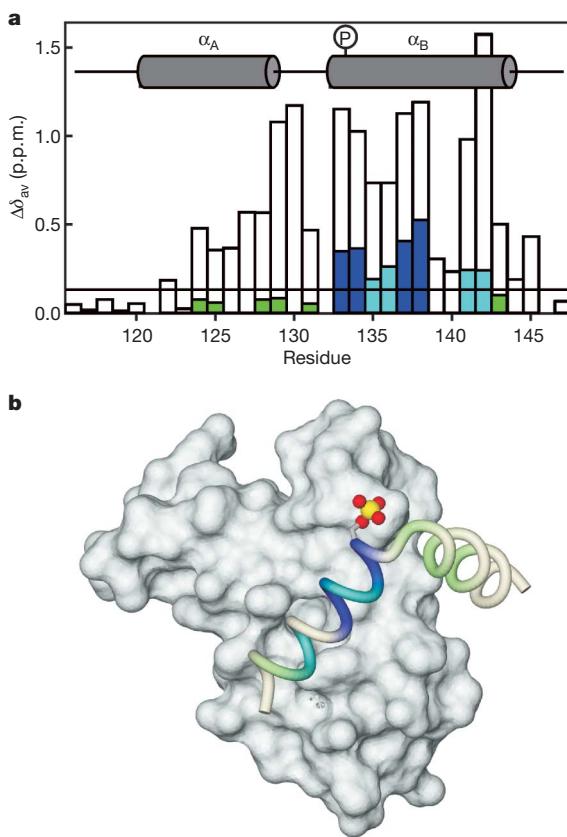


Figure 2 | Characteristics of the encounter complex. **a**, Chemical shift differences for pKID backbone amide (^1H and ^{15}N) on binding to KIX, calculated according to the formula $\Delta\delta_{\text{av}} = \sqrt{(\Delta\delta_{\text{HN}})^2 + (\Delta\delta_{\text{N}}/5)^2}$, where $\Delta\delta_{\text{HN}}$ and $\Delta\delta_{\text{N}}$ are the amide proton and nitrogen chemical shift differences, respectively. The horizontal line corresponds to the average chemical shift difference (0.131 p.p.m.) between the free and the fast-exchanging states. Blue bars indicate residues that exhibit chemical shift deviations exceeding the average by more than 1 s.d. (0.289 p.p.m.), and cyan and pale green bars denote residues with intermediate and below average shift differences, respectively. The open bars indicate chemical shift differences between free pKID and fully bound pKID. The locations of the two helices α_A and α_B of pKID and the phosphorylation site are shown at the top of the figure. **b**, Structure of the high-affinity pKID:KIX complex¹¹, showing a backbone representation of pKID bound to the KIX surface. Chemical shift differences between the free and fast-exchanging states are mapped onto the pKID backbone using the same colour scheme as in **a**. The location of the pSer 133 side chain is shown. The figure was prepared using MOLMOL²⁹.

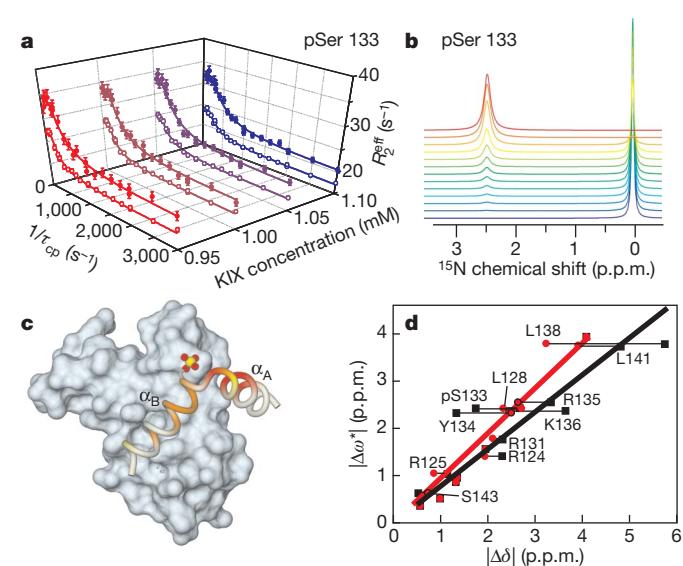


Figure 3 | Analysis of R_2 dispersion experiments. **a**, ^{15}N R_2 relaxation dispersion profile for pSer 133 of pKID recorded at 800 MHz (filled circles) and 500 MHz (open circles). Dispersion curves for 1 mM [^{15}N]-pKID in the presence of 0.95, 1.00, 1.05 and 1.10 mM KIX are shown. **b**, Simulated ^{15}N NMR spectra of pSer 133 at a ^{15}N resonance frequency of 81.1 MHz. The spectra were simulated using the parameters obtained by fitting the R_2 dispersion curves to the three-site exchange model. Chemical shifts in the free form are set to zero. The pKID:KIX concentration ratios are depicted using the same colour scheme as Fig. 1. **c**, Extent of folding in the intermediate state as indicated by $\Delta\omega_{\text{FI}}/\Delta\omega^*$ ratios, displayed on the structure of the high-affinity complex. Red, $\Delta\omega_{\text{FI}}/\Delta\omega^* > 0.9$; orange, $0.5 < \Delta\omega_{\text{FI}}/\Delta\omega^* < 0.9$; yellow, $\Delta\omega_{\text{FI}}/\Delta\omega^* < 0.5$. **d**, Correlation of ^{15}N chemical shift differences ($\Delta\omega^*$) determined from the R_2 dispersion measurements with equilibrium shift differences ($\Delta\delta$). Chemical shift differences between free pKID and the fully bound state ($\Delta\delta_{\text{FB}}$) are shown as black squares, and between the encounter complex and fully bound state ($\Delta\delta_{\text{EB}}$) are shown as red circles, with matching colours for the lines of best fit. The scatter plots show a better correlation between $\Delta\omega^*$ and $\Delta\delta_{\text{EB}}$ (slope = 0.95, $R^2 = 0.94$) than between $\Delta\omega^*$ and $\Delta\delta_{\text{FB}}$ (slope = 0.79, $R^2 = 0.79$). This difference was shown to be statistically significant using an F-test. The simulated and experimental chemical shift differences were obtained from the simulated ^{15}N NMR spectra (**b**) and from ^1H - ^{15}N HSQC spectra (Fig. 1), respectively. $\Delta\delta_{\text{EB}}$ was calculated using the estimated fully bound chemical shifts of the encounter complex (magenta circles in Fig. 1). For resonances that undergo no fast-exchange shifts, the corresponding $\Delta\delta_{\text{FB}}$ values are used.

Table 1 | R_2 dispersion curve fitting with a three-site exchange model

Cluster	Residue	$\Delta\omega_{\text{FI}}^*$ (p.p.m.)	$\Delta\omega_{\text{FB}}^*$ (p.p.m.)	$k_{\text{on}}^* \times 10^6$ ($\text{M}^{-1} \text{s}^{-1}$)	k_{off}^* (s^{-1})	Apparent k_{off} (s^{-1})	k_{IB} (s^{-1})	k_{BI} (s^{-1})	$K_D^* $ (μM)
I	Arg 124	1.33 (0.02) [‡]	1.71 (0.02) [‡]	6.5 (0.1) [‡]	13.2 (0.6) [‡]	9.3	394 (8) [‡]	928 (10) [‡]	1.4
	Arg 125	1.17 (0.03)	0.84 (0.03)	5.0 (0.2)	15 (1)	10.4			2.1
	Ile 127	1.47 (0.03)	1.87 (0.03)	7.3 (0.2)	13 (1)	9.0			1.2
	Leu 128	2.31 (0.03)	2.79 (0.03)	8.1 (0.2)	9.9 (0.7)	6.9			0.9
II	Arg 130	2.12 (0.11)	4.01 (0.06)	5.4 (0.3)	155 (21)	5.2	2241 (25)	78 (6)	1.0
	Arg 131	1.77 (0.03)	3.25 (0.02)	15 (1)	4.5 (0.3)	4.4	34 (2)	957 (7)	0.3
II	pSer 133	1.76 (0.04)	2.53 (0.04)	5.5 (0.2)	54 (6)	6.2	1718 (10)	224 (10)	1.1
	Tyr 134	1.72 (0.05)	2.44 (0.06)	6.4 (0.3)	56 (7)	6.5			1.0
	Arg 135	1.89 (0.05)	2.66 (0.05)	8.9 (0.3)	65 (6)	7.5			0.9
	Lys 136	1.66 (0.03)	2.48 (0.03)	7.1 (0.2)	55 (4)	6.4			0.9
	Leu 138	3.12 (0.05)	3.90 (0.05)	7.3 (0.3)	48 (5)	5.6			0.8
	Asn 139	1.22 (0.05)	0.89 (0.03)	4.2 (0.3)	137 (14)	15.8			3.8
III	Asp 140	0.6 (0.1)	0.59 (0.01)	5.5 (0.5)	601 (43)	61.6	2852 (16)	326 (18)	11.3
	Leu 141	2.73 (0.05)	3.88 (0.06)	6.3 (0.3)	69 (6)	7.1			1.1
	Ser 143	0.92 (0.06)	0.63 (0.03)	4.3 (0.2)	89 (12)	9.1			2.1
	Asp 144	0.19 (0.06)	0.43 (0.02)	5.9 (0.4)	213 (33)	21.9			3.7
IV	Ala 145	1.11 (0.01)	0.81 (0.01)	7.62 (0.05)	16.9 (0.6)	10.1	340 (5)	511 (7)	1.3

See text for definitions of quantities in column headings.

† Chemical shift differences $\Delta\omega_{\text{FI}}$ and $\Delta\omega_{\text{FB}}$ (in rad s^{-1}) were used in curve fitting; values shown here, in units of p.p.m., are calculated according to the formula: $\Delta\omega(\text{p.p.m.}) = \Delta\omega(\text{rad s}^{-1})/(2\pi B_0)$, where B_0 is the ^{15}N resonance frequency.

‡ Values in parentheses are standard deviations.

§ Apparent k_{off} are calculated as described in Methods.

|| Effective dissociation constants are calculated as described in Methods. Excluding the outlier Asp 140, the mean K_D^* obtained by fitting the dispersion data with a three-site exchange model is $1.5 \pm 1.0 \mu\text{M}$.

METHODS SUMMARY

^{15}N R_2 relaxation rates were measured for $[^{15}\text{N}]\text{-pKID:KIX}$ at 1:0.95, 1:1.0, 1:1.05 and 1:1.10 concentration ratios using relaxation-compensated constant-time Carr-Purcell-Meiboom-Gill pulse sequences^{15,27}. Dispersion curves at all four pKID:KIX concentration ratios and at two spectrometer frequencies were fitted simultaneously for each residue using the program GLOVE²⁸. Data were fitted to two-site exchange ($\text{F} \leftrightarrow \text{B}$) and three-site exchange ($\text{F} \leftrightarrow \text{I} \leftrightarrow \text{B}$) models. The fitting parameters are a population-average intrinsic relaxation rate R_2^0 , effective association and dissociation rate constants k_{on}^* and k_{off}^* , folding and unfolding rate constants k_{IB} and k_{BI} , chemical shift differences between each pair of states $\Delta\omega_{\text{FI}}$ and $\Delta\omega_{\text{FB}}$, and total protein concentrations [$\text{pKID}]_0$ and $[\text{KIX}]_0$. The parameters k_{IB} , k_{BI} and $\Delta\omega_{\text{FI}}$ apply only to three-site exchange. The rate constant for the binding process ($\text{F} \rightarrow \text{B}$ for two-site exchange and $\text{F} \rightarrow \text{I}$ for three-site exchange) depends on the concentration of free KIX, which is obtained from the effective dissociation constant K_D^* , determined independently for each residue by fitting the dispersion data, and the total concentrations of pKID and KIX:

$$[\text{KIX}] = \frac{1}{2} \left\{ -K_D^* + a[\text{KIX}]_0 - [\text{pKID}]_0 + \sqrt{(K_D^* - a[\text{KIX}]_0 + [\text{pKID}]_0)^2 + 4a[\text{KIX}]_0 K_D^*} \right\},$$

where a is the pKID:KIX concentration ratio ($a = 0.95, 1, 1.05$ or 1.1). The dispersion data at the different concentration ratios are related to each other through this equation. For the two-site exchange model, $K_D^* = k_{\text{off}}^*/k_{\text{on}}^*$, while for three-site exchange, $K_D^* = k_{\text{off}}^*/k_{\text{on}}^* \times k_{\text{BI}}/(k_{\text{BI}} + k_{\text{IB}})$, in which k_{off}^* is modified by the rate constants k_{BI} and k_{IB} (apparent $k_{\text{off}} = k_{\text{off}}^* \times k_{\text{BI}}/(k_{\text{BI}} + k_{\text{IB}})$). In the three-site exchange model, the intermediate is also a bound state and pKID can dissociate from KIX only from the intermediate state. Therefore, k_{off}^* is scaled by the population ratio: $p_I/(p_I + p_B) = k_{\text{BI}}/(k_{\text{BI}} + k_{\text{IB}})$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 December 2006; accepted 18 April 2007.

Published online 23 May 2007.

1. Tsai, C. J., Kumar, S., Ma, B. Y. & Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci.* **8**, 1181–1190 (1999).
2. Levy, Y., Cho, S. S., Onuchic, J. N. & Wolynes, P. G. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.* **346**, 1121–1145 (2005).
3. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
4. Dyson, H. J. & Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60 (2002).
5. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature Rev. Mol. Cell Biol.* **6**, 197–208 (2005).

6. Uversky, V. N. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* **11**, 739–756 (2002).
7. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc. Natl Acad. Sci. USA* **97**, 8868–8873 (2000).
8. Chiriva, J. C. et al. Phosphorylated CREB binds specifically to nuclear protein CBP. *Nature* **365**, 855–859 (1993).
9. Kwok, R. P. S. et al. Nuclear protein CBP is a coactivator for the transcription factor CREB. *Nature* **370**, 223–226 (1994).
10. Parker, D. et al. Phosphorylation of CREB at Ser133 induces complex formation with CBP via a direct mechanism. *Mol. Cell. Biol.* **16**, 694–703 (1996).
11. Radhakrishnan, I. et al. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions. *Cell* **91**, 741–752 (1997).
12. Shih, H. M. et al. A positive genetic selection for disrupting protein-protein interactions: Identification of CREB mutations that prevent association with the coactivator CBP. *Proc. Natl Acad. Sci. USA* **93**, 13896–13901 (1996).
13. Shaywitz, A. J., Dove, S. L., Kornhauser, J. M., Hochschild, A. & Greenberg, M. E. Magnitude of the CREB-dependent transcriptional response is determined by the strength of the interaction between the kinase-inducible domain of CREB and the KIX domain of CREB-binding protein. *Mol. Cell. Biol.* **20**, 9409–9422 (2000).
14. Parker, D. et al. Analysis of an activator:coactivator complex reveals an essential role for secondary structure in transcriptional activation. *Mol. Cell* **2**, 353–359 (1998).
15. Loria, J. P., Rance, M. & Palmer, A. G. A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.* **121**, 2331–2332 (1999).
16. Mulder, F. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. Studying excited states of proteins by NMR spectroscopy. *Nature Struct. Biol.* **8**, 932–935 (2001).
17. Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J. Biol. Chem.* **277**, 43168–43174 (2002).
18. De Guzman, R. N., Goto, N. K., Dyson, H. J. & Wright, P. E. Structural basis for cooperative transcription factor binding to the CBP coactivator. *J. Mol. Biol.* **355**, 1005–1013 (2006).
19. Gabdoulline, R. R. & Wade, R. C. Biomolecular diffusional association. *Curr. Opin. Struct. Biol.* **12**, 204–213 (2002).
20. Berg, O. G. & von Hippel, P. H. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem.* **14**, 131–160 (1985).
21. Tang, C., Iwahara, J. & Clore, G. M. Visualization of transient encounter complexes in protein-protein association. *Nature* **444**, 383–386 (2006).
22. Volkov, A. N., Worrall, J. A. R., Holtzmann, E. & Ubbink, M. Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase

determined by paramagnetic NMR. *Proc. Natl Acad. Sci. USA* **103**, 18945–18950 (2006).

- 23. Zor, T., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol.* **337**, 521–534 (2004).
- 24. Johnson, C. S. Jr & Moreland, C. G. The calculation of NMR spectra for many-site exchange problems. *J. Chem. Ed.* **50**, 477–483 (1973).
- 25. Shoup, D. & Szabo, A. Role of diffusion in ligand binding to macromolecules and cell-bound receptors. *Biophys. J.* **40**, 33–39 (1982).
- 26. Selzer, T. & Schreiber, G. New insights into the mechanism of protein-protein association. *Proteins Struct. Funct. Genet.* **45**, 190–198 (2001).
- 27. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001).
- 28. McElheny, D., Schnell, J. R., Lansing, J. C., Dyson, H. J. & Wright, P. E. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc. Natl Acad. Sci. USA* **102**, 5032–5037 (2005).
- 29. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Martinez-Yamout for assistance with NMR titrations. This work was supported by the National Institutes of Health and the Skaggs Institute for Chemical Biology.

Author Contributions K.S. and P.E.W. designed the experiments, and K.S., P.E.W. and H.J.D. analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.E.W. (wright@scripps.edu).

METHODS

Sample preparation. Uniformly ^{15}N -labelled KID domain (residues 116–147) of rat CREB was expressed as a GB1 fusion protein³⁰ in BL21-DE3 cells in M9 minimal medium; Ser 133 phosphorylation was accomplished in *Escherichia coli* by coexpression with protein kinase A. The GB1 fusion protein was cleaved with Factor Xa and pKID was purified by reverse-phase HPLC. Unlabelled KIX domain (residues 586–672) of mouse CBP was prepared as described¹¹. The proteins were dissolved separately in NMR buffer (90% $\text{H}_2\text{O}/10\%$ D_2O , 20 mM Tris-d₁-acetate-d₄ (pH 7.0 at 30 °C), 50 mM NaCl, 2 mM NaN₃) and concentrated. Samples of the [^{15}N]-pKID:KIX complex for the R_2 dispersion experiments were prepared from a single concentrated solution of each protein to make the concentration ratios accurate; the pKID concentration was kept at 1.0 mM while KIX concentration was 0.95, 1.0, 1.05 or 1.1 mM. The concentration was determined from the absorbance at 280 nm, using extinction coefficients of 12.95 mM⁻¹ cm⁻¹ for KIX and 1.49 mM⁻¹ cm⁻¹ for pKID.

NMR methods. ^{15}N R_2 relaxation rates were measured on Bruker 500 and 800 MHz spectrometers at a temperature of 30 °C using relaxation-compensated constant-time Carr-Purcell-Meiboom-Gill (CPMG) pulse sequences^{15,27}. R_2 dispersion spectra were acquired as two-dimensional data sets with a constant relaxation delay of 40 or 60 ms. Four data points, including a reference spectrum acquired with the CPMG blocks omitted, were collected in duplicate and were used to estimate the absolute uncertainties and the signal-to-noise ratio of each spectrum. The K_D (680 μM) for the fast exchanging complex was estimated by fitting the shifts of KIX resonances induced by excess pKID.

Fitting of R_2 dispersion profiles. Dispersion curves were fitted to two-site exchange ($\text{F} \leftrightarrow \text{B}$) and three-site exchange ($\text{F} \leftrightarrow \text{I} \leftrightarrow \text{B}$) models using the in-house program GLOVE²⁸. The fitting parameters are a population-average intrinsic relaxation rate R_2^0 , effective association and dissociation rate constants k_{on}^* and k_{off}^* , folding and unfolding rate constants k_{IB} and k_{BI} , chemical shift differences between each pair of states $\Delta\omega_{\text{FI}}$ and $\Delta\omega_{\text{FB}}$, and total protein concentrations [pKID]₀ and [KIX]₀. The fitting parameters k_{IB} , k_{BI} and $\Delta\omega_{\text{FI}}$ apply only to the three-site exchange model. If the kinetic rate constants are much faster than the differences in intrinsic relaxation rate between the states, the intrinsic relaxation rate for each state cannot be derived³¹. The rate constant for the binding process ($\text{F} \rightarrow \text{B}$ for the two-site exchange model and $\text{F} \rightarrow \text{I}$ for the three-site exchange model) depends on the concentration of free KIX, which is obtained from the effective dissociation constant K_D^* , determined independently for each residue by fitting the dispersion data, and the total concentrations of pKID and KIX:

$$[\text{KIX}] = \frac{1}{2} \left\{ -K_D^* + a[\text{KIX}]_0 - [\text{pKID}]_0 + \sqrt{(K_D^* - a[\text{KIX}]_0 + [\text{pKID}]_0)^2 + 4a[\text{KIX}]_0 K_D^*} \right\} \quad (1)$$

where a is the pKID:KIX concentration ratio ($a = 0.95, 1, 1.05$ or 1.1). For each residue, the dispersion data at the different concentration ratios are related to each other through this equation. For the two-site exchange model, $K_D^* = k_{\text{off}}^*/k_{\text{on}}^*$, while for three-site exchange, $K_D^* = k_{\text{off}}^*/k_{\text{on}}^* \times k_{\text{BI}}/(k_{\text{BI}} + k_{\text{IB}})$, in which k_{off}^* is modified by the rate constants k_{BI} and k_{IB} (apparent $k_{\text{off}} = k_{\text{off}}^* \times k_{\text{BI}}/(k_{\text{BI}} + k_{\text{IB}})$). In the three-site exchange model, the intermediate state is also a bound form and pKID can dissociate from KIX only from the intermediate state. Therefore, k_{off}^* must be scaled by the population ratio: $p_I/(p_I + p_B) = k_{\text{BI}}/(k_{\text{BI}} + k_{\text{IB}})$.

The R_2 dispersion curves at all four pKID:KIX concentration ratios and at the two spectrometer frequencies were fitted simultaneously for each residue. Fits were initially performed for each individual residue and the goodness of fit was assessed by the reduced χ^2 value (χ^2 divided by the degree of freedom, designated $\chi^2_{\text{r.individual}}$). However, it is reasonable to assume that the folding and unfolding processes for residues in a local element of secondary structure will be correlated; neighbouring residues within a given structural element are likely to fold and unfold at similar rates. The R_2 dispersion curves of pKID were therefore fitted in clusters; the folding and unfolding rate constants k_{IB} and k_{BI} were treated as global parameters for all residues in the cluster and the goodness of fit was assessed from $\chi^2_{\text{r.cluster}}$ values. The two pKID helices were treated separately in the cluster analysis. The initial cluster in each helix consisted of two neighbouring residues; additional residues were added to extend each cluster in subsequent rounds of data fitting. For each cluster, the goodness of fit was assessed by comparing the reduced χ^2 value to that obtained by fitting the constituent residues individually. The fitting quality does not change for a correctly identified cluster, but the reduced χ^2 value decreases with the global parameters because the degree of freedom increases. The cluster analysis was therefore performed by increasing the size of each cluster one residue at a time until the ratio of the reduced χ^2 values ($\chi^2_{\text{r.cluster}}/\chi^2_{\text{r.individual}}$) increased and became greater than 1. At this point, the added residue was omitted and new clusters were tested. Residues that lie outside the pKID helices were not included in the cluster analysis but their dispersion curves were fitted individually.

Exchange models. For the two-site exchange model, the following analytical equation³² can be used:

$$R_2^{\text{eff}} = R_2^0 + \frac{1}{2} \left\{ [\text{KIX}] k_{\text{on}}^* + k_{\text{off}}^* - \frac{1}{\tau_{\text{cp}}} \cosh^{-1} [D_+ \cosh(\eta_+) - D_- \cosh(\eta_-)] \right\} \quad (2)$$

where $D_{\pm} = \frac{1}{2} \left[\pm 1 + \frac{\Psi + 2\Delta\omega_{\text{FB}}^2}{\sqrt{\Psi^2 + \zeta^2}} \right]$,

$$\eta_{\pm} = \tau_{\text{cp}} \sqrt{\frac{1}{2} \left(\pm \Psi + \sqrt{\Psi^2 + \zeta^2} \right)},$$

$$\Psi = ([\text{KIX}] k_{\text{on}}^* + k_{\text{off}}^*)^2 - 4\Delta\omega_{\text{FB}}^2 + 4[\text{KIX}] k_{\text{on}}^* k_{\text{off}}^*,$$

$$\zeta = 2\Delta\omega_{\text{FB}} ([\text{KIX}] k_{\text{on}}^* - k_{\text{off}}^*)$$

where τ_{cp} is the interval between 180° pulses in CPMG. In the case of a three-site exchange model, R_2^{eff} can be calculated according to³³:

$$R_2^{\text{eff}} = R_2^0 - \frac{1}{\tau_{\text{cp}}} \ln(\lambda_1) \quad (3)$$

Under the experimentally accessible condition, R_2^{eff} is dominated by the largest eigenvalue λ_1 of the matrix

$$\begin{pmatrix} R[\exp(-\mathbf{A} \frac{\tau_{\text{cp}}}{2}) \exp(-\mathbf{A}^* \frac{\tau_{\text{cp}}}{2})] & I[\exp(-\mathbf{A} \frac{\tau_{\text{cp}}}{2}) \exp(-\mathbf{A}^* \frac{\tau_{\text{cp}}}{2})] \\ I[\exp(-\mathbf{A} \frac{\tau_{\text{cp}}}{2}) \exp(-\mathbf{A}^* \frac{\tau_{\text{cp}}}{2})] & -R[\exp(-\mathbf{A} \frac{\tau_{\text{cp}}}{2}) \exp(-\mathbf{A}^* \frac{\tau_{\text{cp}}}{2})] \end{pmatrix} \quad (4)$$

where $R[\cdot]$ and $I[\cdot]$ are functions to extract the real or imaginary elements, respectively, of the complex matrix. As the matrix \mathbf{A} is a 3-by-3 evolution matrix as given below (\mathbf{A}^* is its complex conjugate) for the three-site exchange model, the matrix size shown as (4) is 6-by-6. If the kinetic rate constants are much faster than differences in intrinsic relaxation rate between the sites,

$$\mathbf{A} = \begin{pmatrix} [\text{KIX}] k_{\text{on}}^* & -k_{\text{off}} & 0 \\ -[\text{KIX}] k_{\text{on}}^* & k_{\text{off}} + k_{\text{IB}} - i\Delta\omega_{\text{FI}} & -k_{\text{BI}} \\ 0 & -k_{\text{IB}} & k_{\text{BI}} - i\Delta\omega_{\text{FB}} \end{pmatrix} \quad (5)$$

Effective dissociation constant for the three-site exchange model. In our model, the intermediate state is one of the bound forms. Therefore, the effective dissociation constant K_D^* is given by

$$K_D^* = \frac{[\text{pKID}][\text{KIX}]}{[\text{pKID} : \text{KIX}]_1 + [\text{pKID} : \text{KIX}]_B} \quad (6)$$

where $[\text{pKID}]$, $[\text{KIX}]$, $[\text{pKID} : \text{KIX}]_1$ and $[\text{pKID} : \text{KIX}]_B$ are the concentrations of free pKID, free KIX, complex in the intermediate state, and complex in the fully bound state, respectively. When both numerator and denominator are divided by the total pKID concentration, K_D^* can be expressed by the populations of the free, intermediate and bound states together with [KIX],

$$K_D^* = \frac{p_F[\text{KIX}]}{p_I + p_B} \quad (7)$$

All populations can be derived from the kinetic rate constants using the condition of microscopic reversibility,

$$p_F = \frac{k_{\text{off}}^* k_{\text{BI}}}{k_{\text{off}}^* k_{\text{BI}} + k_{\text{on}}^* [\text{KIX}] (k_{\text{IB}} + k_{\text{BI}})} \quad (8)$$

$$p_I = \frac{k_{\text{on}}^* [\text{KIX}] k_{\text{BI}}}{k_{\text{off}}^* k_{\text{BI}} + k_{\text{on}}^* [\text{KIX}] (k_{\text{IB}} + k_{\text{BI}})}$$

$$p_B = \frac{k_{\text{on}}^* [\text{KIX}] k_{\text{IB}}}{k_{\text{off}}^* k_{\text{BI}} + k_{\text{on}}^* [\text{KIX}] (k_{\text{IB}} + k_{\text{BI}})}$$

Substitution of equation (8) into (7) gives

$$K_D^* = \frac{k_{\text{off}}^*}{k_{\text{on}}^*} \times \frac{k_{\text{BI}}}{k_{\text{BI}} + k_{\text{IB}}} \quad (9)$$

Calculation of NMR spectra for multiple-site exchange. Excited magnetization at each resonance frequency $M(\omega)$ in the presence of multiple-site exchange can be calculated by²⁴:

$$M(\omega) = -\omega M_0 R[\mathbf{P} \cdot (\mathbf{A}^t)^{-1} \cdot \mathbf{1}] \quad (10)$$

where M_0 is the equilibrium magnetization, which only scales the calculated NMR spectrum and therefore an arbitrary value can be used unless noise is taken into account. $R[\cdot]$ is the function to extract the real part of the complex number.

\mathbf{P} is a row vector whose elements are the population in each state. \mathbf{A} is the Kubo-Sack matrix²⁴. $\mathbf{1}$ is a column vector whose elements are all 1. In the case of our three site-exchange model,

$$\mathbf{P} = (p_F \quad p_I \quad p_B)$$

$$\mathbf{A}(\omega) = \begin{pmatrix} -R_{2F}^0 - [KIX]k_{on}^* + i\omega & k_{off}^* & 0 \\ [KIX]k_{on}^* & -R_{2B}^0 - k_{off}^* - k_{BI} - i(\Delta\omega_{FI} - \omega) & k_{BI} \\ 0 & k_{BI} & -R_{2B}^0 - k_{BI} - i(\Delta\omega_{FB} - \omega) \end{pmatrix} \quad (11)$$

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

The free KIX concentration at each pKID:KIX concentration ratio used in the dispersion measurements was calculated by equation (1) and populations were then derived using equation (8). The intrinsic relaxation rates in the free form (R_{2F}^0) were experimentally determined by measuring ^{15}N R_2 dispersions with pKID in the free form. The R_2^0 rates obtained with the 1:1 concentration ratio sample were used as R_{2B}^0 .

30. Huth, J. R. et al. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci.* **6**, 2359–2364 (1997).
31. Grey, M. J., Wang, C. & Palmer, A. G. III. Disulfide bond isomerization in basic pancreatic trypsin inhibitor: Multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling. *J. Am. Chem. Soc.* **125**, 14324–14335 (2003).
32. Carver, J. P. & Richards, R. E. General 2-site solution for chemical exchange produced dependence of T2 upon Carr-Purcell pulse separation. *J. Magn. Reson.* **6**, 89–105 (1972).
33. Allerhand, A. & Gutowsky, H. S. Spin-echo studies of chemical exchange. II. Closed formulas for two sites. *J. Chem. Phys.* **42**, 1587–1599 (1965).

naturejobs

JOBS OF THE WEEK

These days, going green is all the rage. Universities, often the very bastions of green thinking and research, have begun to take a look in the mirror and scrutinize their own practices, including those in their science laboratories. Not surprisingly, most could stand to waste less and get greener. Indeed, responsible scientists should start thinking about how their daily work consumes energy, and what they can do to be more energy-efficient.

This isn't necessarily a simple task. Principal investigators have plenty on their plate already, with research, grant applications, hiring and, in some cases, teaching duties. And now they have to shrink their carbon footprint. What, my little lab?

It turns out, though, that labs waste plenty. A traditional fume hood uses as much energy in a year as three US households, as a recent *Nature* article points out (see *Nature* **445**, 590–591; 2007). The US Department of Energy's Fermilab in Batavia, Illinois spends \$1 million per month on electricity.

Several groups have taken notice. In Britain, a student group called People and Planet ranked the greenest UK campuses and published the list in *The Times Higher Education Supplement* this month. The group rated factors such as a 'green travel plan', an institution's efforts to organize transport initiatives, and their success at retrieving energy from renewable sources. Leeds Metropolitan University came top.

In the United States, a government-sponsored association, Labs21, helps advise scientists on efficient laboratory design and equipment. In December of last year, American university and college presidents announced a Climate Commitment. Institutions that sign up promise to initiate policies to combat climate change. They're expected, for example, to compile, within one year of signing, an inventory of all of their school's greenhouse-gas emissions from electricity, heating, commuting and air travel. The pledge has 280 signatories so far.

Scientists and their laboratories are not the major culprits in un-green practices. But it makes sense that they're part of the solution — which means adding 'getting green' to an already long list of daily tasks.

Gene Russo, acting editor, Naturejobs

CONTACTS

Acting Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelia Pequignot (4974)
e-mail: a.pequignot@nature.com
Naturevents:
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Naturejobs online production:

Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless
Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

MOVERS

Andreas Meyer-Lindenberg, director, Central Institute of Mental Health, Mannheim, Germany



2005-07: Chief, Unit for Systems Neuroscience in Psychiatry, National Institute of Mental Health (NIMH), Bethesda, Maryland
2004-07: Co-director, Neuroimaging Facility, NIMH
2001-05: Staff clinician, Clinical Brain Disorders Branch (CBDB), NIMH
1997-2001: Visiting associate researcher fellow, CBDB

Andreas Meyer-Lindenberg owes his interest in neuroscience to his psychiatrist father. And now his work could help revolutionize his father's field by offering better options for diagnosis and treatment of mental illness.

At the medical school of the University of Bonn in Germany, Meyer-Lindenberg initially focused on neurochemistry: he did a thesis on receptor mechanisms associated with risk of suicide. Then switched gears to investigate mental illness by understanding how the brain is wired. After a neurology residency, he moved to the Justus-Liebig University hospital in Giessen to conduct neuroimaging research. This "failed miserably", he says, in part because the technology was relatively new and no local partners were interested in applying it to brain study.

Meyer-Lindenberg chose to pursue a postdoc in the United States at the National Institute of Mental Health. Intending to stay for two years to learn neuroimaging methods, he has stayed for ten: in recent years, for example, he has investigated the interaction of the prefrontal cortex and striatum in people with schizophrenia. He also studied mathematical models, earning a master's degree in mathematics from the University of Hagen in Germany. "I was hoping to get a broader tool belt to look at these things," he says.

His mathematical skills have helped him develop methods to investigate complex interactions between genetic variants and their influence on the human brain. He now combines studies of genetic indicators of mental illness with neuroimaging. The approach helps uncover the elusive biological mechanisms of mental disorders, he says.

Starting on 1 July he plans to further that aim in his new role as director of Germany's Central Institute of Mental Health. His predecessor, Fritz Henn, has no doubt he will succeed. "Andreas is the brightest young guy in the field," Henn says, adding that Meyer-Lindenberg pioneered the marriage of genetics and imaging in neurobiological research. Henn has spent the past five years trying to recruit Meyer-Lindenberg — all the while helping the institute excel in biological psychiatry and making it proficient in both genetics and imaging.

His tactic worked. Meyer-Lindenberg says that the institute's top-notch facilities and departments will allow his group to work towards a new level of therapeutic application. His long-term goal, he adds, is to overhaul how mental illness is categorized and treated. ■

Virginia Gewin

NETWORKS & SUPPORT

Employing Japan's postdocs

Osaka University in Japan recently surveyed 83 of its science and technology postdocs and found that more than 60% no longer wanted to be postdocs. More than 85% felt insecure about their careers. Why are so many PhD recipients at Osaka concerned about their future?

In the early 1990s, the Japanese government attempted to strengthen the country's graduate schools. As a result, the number of PhDs awarded each year increased from 8,968 in 1996 to 15,966 in 2004. However, many PhD recipients fail to get permanent jobs.

The government has tried to help. In 1996 a programme was launched to support 10,000 postdocs by 2000. Meanwhile, Japan's science and technology budget remains high, totalling over 3 trillion yen (about US\$24.5 billion at today's rates) per year since 2001.

Despite this investment, job security is hard to find. Any small growth in the number of permanent academic positions in the past ten years has been overwhelmed by the growth in the number of PhDs awarded. Only a decade ago, most PhD recipients found permanent positions. Today, some PhDs have taken as many as four different three-year jobs by the age of 40.

The Japanese government chose Osaka University as one of eight organizations to take part in a 2006 project to promote the diversification of career paths for science and technology researchers. Some organizations support postdocs by educating them about the business world or providing individual consultation. However, we think that it's not only important to find people jobs, but to create new jobs and new types of job.

More opportunities must be created in a variety of fields. Universities can help solve the problem, as they receive a budget from the government that they can use to create projects and so also jobs. Cooperation between industry and the academic world also creates jobs such as project manager, technology coordinator and programme officer.

At Osaka's career-creation support division, we hope that our activities help increase the number of postdocs who consider unconventional career paths — and, as a result, make them feel more secure about their futures. ■

Yasuo Kanamatsu and Kyoko Takahashi are in the career-creation support division of Osaka University's Center for Advanced Science and Innovation.

POSTDOC JOURNAL

Growth

My mom always says that we should grow where we're planted. I take this advice to heart, especially since my parents flourished after immigrating to the United States with four small children, 12 big boxes and not much else. As a postdoc, I live by these words, particularly since my husband and I relocated to London to pursue our research fellowships, albeit without children, and with many more boxes, but with a similar challenge of acclimatizing to a foreign land.

Scientific mobility often means successive moves to different environments to which we must adapt, and in which we must thrive, in order to be productive scientists. Naturally we seek places where our development will be nurtured, both professionally and personally. Nonetheless, we occasionally face inclement conditions — such as a drought in funding or an unrelenting deluge of responsibilities — that we must endure and survive in order to be fruitful.

However, it is in such climates that we can discover the depths of our own resources. Deeply rooted values and goals can provide us with the tenacity to hold our ground and to weather various storms, thus allowing us to bloom even in the toughest of fields. So the next time I feel buried under a pile of manure, I'll see it as an opportunity to cultivate my strengths and to grow. After all, it's just more fertilizer. ■

Maria Thelma Ocampo-Hafalla is a research fellow at Cancer Research UK's London Research Institute.

The inside track from academia and industry

Time to make haste

Scientists with a busy schedule must free up time to develop professionally in a fast-moving sector.



Deb Koen

Scientists often don't know the best way to incorporate professional development into already busy research schedules. It's a theme that frequently emerges in my presentations and conversations with the research community. In fact, scientists can't afford *not* to carve out at least a portion of their time for career development. Those who decline are not only standing still but losing ground. With precious little warning, a skill set you're wedded to can become obsolete, while new priorities arise almost overnight. Whether you live in the lab, carry a full teaching course load or manage a research division, it's important to make development a priority. Translating intention into action is the key. Here's my top-ten checklist of success strategies.

Play to your strengths

Being true to yourself is the only solid foundation on which to build a career. When you consider career development, avoid the tendency to focus only on weaknesses to the exclusion of your strengths. Instead, concentrate your efforts on work that gives you personal satisfaction while leveraging your strong points.

"With little warning, a skill set you're wedded to can become obsolete."

whose opinion you respect, such as a boss, mentor or colleague, for suggestions on how to improve your work, your style and your approach. View their insights as gifts.

Bring someone else along

Develop yourself by developing others. Get a picture in your mind of someone to whom you would like to reach out. Maybe you'll become an active mentor

Seek feedback

Feedback will provide the framework within which to grow. Become aware of blind spots by asking others

or play an advocacy role. Act on opportunities to contribute to the individual's career growth. Whether recommending a course, making a connection to an influential researcher for a mentee or passing along information on openings, your efforts to help another will result in a shared learning experience.

Study your organization

With an all-consuming schedule, it's easy to operate in a vacuum. The danger with having too narrow a focus is that while you're busy working hard, your goals may fall out of line with your organization's. Camera maker Eastman Kodak, for example, had to deal with a revolutionary wave of digital technology: employees who saw it coming tweaked their skill sets and career paths; others found themselves out of a job. Broaden your perspective to allow for changing trends and career advancement.

Target a trend

Expand your study beyond your current research to stay informed of developments in your field, as well as the social, political and economic realms. Governmental regulations in stem-cell research or increased focus on and money for biofuels research, for example, could have an impact on career development choices. It's no secret to scientists that research initiatives fall in and out of favour with various social movements and changes in political administrations. Consider joining the World Future Society (www.wfs.org) for a big picture look at predictions.

Build your brand

Branding involves developing and managing your reputation through your results and your relationships. Track professional accomplishments by keeping a career file, constantly looking for what makes you unique. Communicate your brand

through a CV or resumé, presentations (formal and informal), a website, if you have one, and any other relevant marketing materials.

Adopt a healthful practice

Healthy habits are easily crushed under mounting workplace pressures. With sedentary roles and stressful environments, it's unlikely that your need for exercise, nutrition, spiritual renewal and relaxation will be met without conscious effort. Through meditation, a yoga class or a buddy system, embracing a healthful practice is essential to your overall development.

Pick a new constituency

There's probably a constituency out there that you've been meaning to reach out to. Whether you contact funders, focus on your lab mates or become active in a scientific association, connecting to an overlooked group could enhance performance and contribute to your professional development.

Stretch your boundaries

Specialization is necessary but not sufficient in a world where international connections and contacts are increasingly important. Diversify your relationships. Mingle with new groups. Frequent places you've never been. Read books and journals out of your discipline. In essence, develop by broadening your perspective.

Simplify

Given that researchers are generally extremely busy, it's difficult to focus on all ten strategies here. To maximize your time and energy, figure out what complicates your life and then set out to simplify. In addition to simplifying, select and apply two other strategies that best reflect your development interests.

Deb Koen is the president and CEO of Career Development Services in Rochester, New York.